

Thu May 1 15:54:21 2003

Reun

US-09-446-543a-73.Tag

SEQ ID NO: 73

Database: A-Geneseg-101002

AC NO: AAW31394

File City
Page 1
Printed on
paper #16

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 1, 2003, 14:29:12 ; Search time 30.5938 Seconds
(without alignments)
95.821 Million cell updates/sec

Title: US-09-446-543a-73

Sequence: 1 TPDINPARYXXRGIRVGRFX 22

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/emb1/AA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/emb1/AA2001.DAT:*
22: /SIDS2/gcgdata/geneseq/emb1/AA2002.DAT:*
23: /SIDS2/gcgdata/geneseq/emb1/AA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	20	AAW31394	Human type G prote
2	105	96.3	20	AAW7236	Human type G prote
3	105	96.3	20	AAW10365	Human type G prote
4	105	96.3	20	AAW49294	Human type G prote
5	105	96.3	20	AAW62534	Human type G prote
6	105	96.3	20	AAW90992	Human type G prote
7	105	96.3	21	AAW31395	Human type G prote
8	105	96.3	21	AAW10366	Human type G prote
9	105	96.3	21	AAW62535	Human type G prote
10	105	96.3	22	AAW31396	Human type G prote

11	105	96.3	22	AAW10367	Human oxytocin sec
12	105	96.3	22	AAW62536	Human CRH releasin
13	105	96.3	31	AAW31391	Human type G prote
14	105	96.3	31	AAW97235	Human type G prote
15	105	96.3	31	AAW87615	Human type G prote
16	105	96.3	31	AAW10362	Human type G prote
17	105	96.3	31	AAW49291	Human oxytocin sec
18	105	96.3	31	AAW62531	Human CRH releasin
19	105	96.3	31	AAW90991	Human CRH releasin
20	105	96.3	31	AAW30995	Human type G prote
21	105	96.3	32	AAW31392	Human type G prote
22	105	96.3	32	AAW10363	Human oxytocin sec
23	105	96.3	32	AAW62532	Human CRH releasin
24	105	96.3	33	AAW31393	Human type G prote
25	105	96.3	33	AAW10364	Human type G prote
26	105	96.3	33	AAW62533	Human CRH releasin
27	105	96.3	37	AAW31390	Human type G prote
28	105	96.3	87	AAW97226	Human type G prote
29	105	96.3	87	AAW10361	Human oxytocin sec
30	105	96.3	87	AAW62530	Human CRH releasin
31	104	95.4	20	AAW31387	Human type G prote
32	104	95.4	20	AAW10374	Human type G prote
33	104	95.4	20	AAW97232	Human type G prote
34	104	95.4	20	AAW97234	Human type G prote
35	104	95.4	20	AAW95191	Human type G prote
36	104	95.4	20	AAW95175	Human type G prote
37	104	95.4	20	AAW10350	Human type G prote
38	104	95.4	20	AAW10358	Human type G prote
39	104	95.4	20	AAW49301	Human type G prote
40	104	95.4	20	AAW49302	Human type G prote
41	104	95.4	20	AAW62519	Human type G prote
42	104	95.4	20	AAW62527	Human type G prote
43	104	95.4	20	AAW90994	Human type G prote
44	104	95.4	20	AAW90996	Human type G prote
45	104	95.4	20	AAW49354	Human type G prote

ALIGNMENTS

RESULT 1
AAW31394
ID AAW31394 standard; Peptide: 20 AA.
AC AAW31394;
XX 06-APR-1998 (first entry)
XX
XX Human type G protein-coupled receptor ligand fragment 4.
XX
XX G protein-coupled receptor; ligand binding; pharmaceutical;
XX modulator; pituitary; central nervous system; pancreas; prophylactic;
XX therapeutic agent.
XX
XX Homo sapiens.
XX W09724436-A2.
XX
XX 10-JUL-1997.
XX
XX 26-DEC-1996; 96WO-JP03821.
XX
XX 18-SEP-1996; 96JP-0246573.
XX 28-DEC-1995; 95JP-0343371.
XX 15-MAR-1996; 96JP-0059419.
XX 12-AUG-1996; 96JP-0211805.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX Fujii R, Fukushima S, Habata Y, Hinuma S, Hosoya M,
XX Kawamata Y, Kitada C;
XX WPI, 1997-363672/33.

DR N-PSDB; AAV02431.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 185; 258pp; English.

CC This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 34 to 53 of the
CC sequence represented in AAV01390 and is used in an assay to monitor
CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a
CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
CC trauma, growth hormone secretory disease, hyper- and polyphagia,
CC hyperlipidaemia, hypochlosterolemia, hyperglycaemia,
CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
CC acute myocardial infarction, infertility, spinocerebellar degeneration,
CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
CC oligosacchara. Assays can also be developed to screen compounds which are
CC capable of altering the binding activity of the ligand affecting
CC activation of the G protein-coupled receptor protein.

SO Sequence 20 AA;

Query Match

Best Local Similarity 96.3%; Score 105; DB 18; Length 20;

Hatches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMTXXRGIRPVGRF 20
1 TPDINPAMTXXRGIRPVGRF 20

Db

RESULT 2

AAW97236

AC AAW97236;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
XX G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacosgenesis;
XX menopausal syndrome; euthyroid; hypometabolism; lactation;
XX pituitary adenomatosis; brain tumour; emmenopathy; autoimmune disease;
XX prolactinoma; infertility; impotence; amenorrhea; galactorrhoea;
XX acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
XX Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
XX contraceptive; placental function; choriochorionoma; hydatid mole;
XX abnormal lipometabolism; oxytocia.

OS Homo sapiens.

PN W09858962-AL.

PD 30-DEC-1998.

PE 22-JUN-1998; 98WO-JP02765.

PR 23-JUN-1997; 97JP-0165437.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;

DR WPI; 1999-105614/09.

XX Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 166; 241pp; English.

CC The present sequence represents a human type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypovarianism, gonocyst cacosgenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can be used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmenopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhoea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC chorioncarinoma, hydatid mole, abortion, unfertilized fetus,
CC abnormal saccharometabolism, abnormal lipometabolism or oxytocia.

SO Sequence 20 AA;

Query Match

Best Local Similarity 96.3%; Score 105; DB 20; Length 20;

Hatches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMTXXRGIRPVGRF 20
1 TPDINPAMTXXRGIRPVGRF 20

Db

RESULT 3

AAAB10365

AC AAAB10365;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 35.

XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
XX treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
XX caesarean section; artificial fertilization; galactostasis; goat; pig;
XX veterinary medicine; milk production.

OS Homo sapiens.

PN W0200038704-AL.

PD 06-JUL-2000.

PE 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

XX Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine

ps Disclosure; Page 63; 72pp; Japanese.

xx This invention describes a novel oxytocin secretion-regulating agent

cc which contains a ligand peptide or its salt for the G protein-coupled

cc receptor protein. It is useful in the form of drugs for ameliorating,

cc preventing and treating diseases relating to oxytocin secretion e.g.,

cc weak pains and atonic bleeding, before and after expulsion of placenta,

cc uterine recovery failure, caesarean section, stoppage of artificial

cc fertilization or galactostasis and is also applicable in veterinary

cc medicine for promoting milk production in cow, goat and pig. This

cc sequence represents a human peptide which acts as an oxytocin secretion

cc promoter.

xx

sq Sequence 20 AA:

Query Match 96.3%; Score 105; DB 21; Length 20;

Best Local Similarity 90.0%; Pred. No. 3.2e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 4

AAV9294

ID AAV9294 standard; peptide; 20 AA.

AC AAV9294;

XX

DT 22-FEB-2000 (first entry)

DE 19P2 ligand peptide fragment.

XX

KM Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;

KW pituitary; regulatory mechanism; central nervous system; pancreatic.

XX

OS Homo sapiens.

XX

PI Key Location/Qualifiers

FT Modified-site 20

XX /note="C-terminal amide"

PN W0960112-A1.

XX

PD 25-NOV-1999.

XX

PF 20-MAY-1999; 99WO-JP02650.

XX

PR 21-MAY-1998; 98JP-0140293.

XX

PA (TAKE) TAKEDA CHEM IND LTD.

XX

PI Matsumoto H, Kitada C, Hinuma S;

XX

DR WPI; 2000-039381/03.

XX

PT New monoclonal antibodies, useful in diagnosis, as drugs and in

PT studying diseases related to ligand abnormality

XX

PS Disclosure; Page 26; 73pp; Japanese.

XX

CC The invention provides a monoclonal antibody which has a specific

CC reaction with the part peptide of the C-terminal of 19P2 ligand or its

CC derivative. The antibodies can be used in diagnosis or to treat or

CC prevent diseases associated with abnormality in the pituitary function

CC regulatory mechanism (e.g. promotion of prolactin secretion), central

CC nervous regulatory mechanism, and pancreatic function regulatory

CC mechanism. The antibody-based immunoassay can also be applied in

CC clarifying the physiological functions of the ligand and its derivative.

CC Sequences AAV9290-302 represent peptide fragments of the 19P2 ligand.

XX

BQ Sequence 20 AA;

Query Match 96.3%; Score 105; DB 21; Length 20;

Best Local Similarity 90.0%; Pred. No. 3.2e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 5

AA62534

ID AA62534 standard; peptide; 20 AA.

AC AA62534;

XX

DT 24-AUG-2001 (first entry)

DE Human CRH releasing protein related peptide SEQ ID NO: 35.

XX

KM Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;

KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;

KW Addison's disease; adrenal gland hyperfunction; obesity.

XX

OS Homo sapiens.

XX

PN W0200135984-A1.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000WO-JP08119.

XX

PR 18-NOV-1999; 99JP-0327900.

XX

PR 26-SEP-2000; 2000JP-0297073.

XX

PA (TAKE) TAKEDA CHEM IND LTD.

XX

PI Kitada C, Matsumoto H, Hinuma S;

XX

DR WPI; 2001-35552/37.

XX

PT Use of G protein receptor ligand or peptide for controlling

PT corticotropin releasing hormone secretion

XX

PS Claim 4; Page 75; 90pp; Japanese.

XX

CC The present sequence describes a method of controlling the secretion of

CC corticotrophin releasing hormone (CRH), involving the use of a G protein

CC receptor ligand. This can be used to control the secretion of CRH and is

CC useful as an analgesic or for treating, preventing or ameliorating

CC diseases associated with CRH secretion such as hyperaldosteronism

CC hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's

CC disease (including boredom, nausea, pigmentation, hypogonadism, hair

CC loss, and hypotension), adrenal gland hyperfunction and obesity. The

CC present sequence is a peptide used in the exemplification of the

CC invention.

XX

sq Sequence 20 AA;

Query Match 96.3%; Score 105; DB 22; Length 20;

Best Local Similarity 90.0%; Pred. No. 3.2e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 6

AA690992

ID AAB90992 standard; Peptide; 20 AA.

AC AAB90992;

XX 22-JUN-2001 (first entry)
 XX Prolactin releasing peptide SEQ ID NO:166.
 DE
 XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KM hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200069900-A2.
 PD 23-NOV-2000.
 XX
 PF 17-MAY-2000; 2000KO-US13576.
 XX
 PR 17-MAY-1999; 99US-0134406.
 PR 10-SEP-1999; 99US-0153406.
 PR 15-OCT-1999; 99US-0159783.
 XX
 PA (CONU-) CONNUCHEM INC.
 PI Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibautau K;
 DR WPI; 2001-112059/12.
 XX
 PT Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity
 XX
 PS Disclosure; Page 244; 733pp; English.
 XX
 CC The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity
 CC in vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention.
 CC
 XX
 SQ Sequence 20 AA;
 Query Match 96.3%; Score 105; DB 22; Length 20;
 Best local Similarity 90.0%; Pred. No. 3.2e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TPDINPAMYXXRGIRPVGRF 20
 ||||||||| |||||||||
 Db 1 TPDINPAMYASRGIRPVGRF 20
 ||||||||| |||||||||
 RESULT 7
 AAB31395
 ID AAB31395 standard; Peptide: 21 AA.
 AC AAB31395;
 XX
 XX 06-APR-1998 (first entry)
 DE Human type G protein-coupled receptor ligand fragment 5.
 XX

KM G protein-coupled receptor; ligand binding; pharmaceutical;
 KM modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.
 XX
 OS Homo sapiens.
 XX
 PN WO9724436-A2.
 PD 10-JUL-1997.
 XX
 PF 26-DEC-1996; 96MO-JP03821.
 XX
 PR 18-SEP-1996; 96JP-0246573.
 PR 28-DEC-1995; 95JP-0343371.
 PR 15-MAR-1996; 96JP-0059419.
 PR 12-AUG-1996; 96JP-0211805.
 XX
 PA (TAKEDA) TAKEDA CHEM IND LTD.
 PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M,
 PI Kawamata Y, Kitada C;
 DR WPI; 1997-363672/33.
 DR N-PSDB; AAV02432.
 XX
 PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2; Page 186; 258pp; English.
 XX
 CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 34 to 54 of the
 CC sequence represented in AAB31390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidemia, hypercholesterolemia, hyperglycemia,
 CC hyperprolactinemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neuritis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 CC
 XX
 SQ Sequence 21 AA;
 Query Match 96.3%; Score 105; DB 18; Length 21;
 Best local Similarity 90.0%; Pred. No. 3.4e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TPDINPAMYXXRGIRPVGRF 20
 ||||||||| |||||||||
 Db 1 TPDINPAMYASRGIRPVGRF 20
 ||||||||| |||||||||
 RESULT 8
 AAB10366
 ID AAB10366 standard; peptide: 21 AA.
 AC AAB10366;
 XX
 XX 24-NOV-2000 (first entry)
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 36.
 KW Human oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 XX

XX caesarean section; artificial fertilization; galactostasis; goat; pig;
XX veterinary medicine; milk production.
XX Homo sapiens.
XX WO200038704-A1.
XX 06-JUL-2000.
XX 22-DEC-1999; 99WO-JP07199.
XX 25-DEC-1998; 98JP-0369585.
XX (TAKE) TAKEDA CHEM IND LTD.
XX Matsumoto H, Kitada C, Hinuma S;
XX WPI; 2000-452296/39.
XX DR
XX PT Physiologically-active polypeptide recognized as ligand by G
XX protein-coupled receptor protein, for promoting secretion of oxytocin,
XX as drugs for diseases relating to oxytocin secretion and in veterinary
XX medicine -
XX
XX PS Disclosure; Page 63; 72pp; Japanese.
XX
XX CC This invention describes a novel oxytocin secretion-regulating agent
XX which contains a ligand peptide or its salt for the G protein-coupled
XX receptor protein. It is useful in the form of drugs for ameliorating,
XX preventing and treating diseases relating to oxytocin secretion e.g.,
XX weak pains and atonic bleeding, before and after expulsion of placenta,
XX uterine recovery failure, caesarean section, stoppage of artificial
XX fertilization or galactostasis and is also applicable in veterinary
XX medicine for promoting milk production in cow, goat and pig. This
XX sequence represents a human peptide which acts as an oxytocin secretion
XX promoter.
XX
XX SQ Sequence 21 AA;
XX
XX Query Match 96.3%; Score 105; DB 21; Length 21;
XX Best Local Similarity 90.0%; Pred. No. 3,4e-11;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 TPDIINAWYXXRGIRVGRF 20
XX | | | | | | | | | | | | | | | | | | | | | |
XX 1 TPDIINAWYASRGIRVGRF 20
XX
XX RESULT 9
XX AAG62535
XX ID AAG62535 standard; peptide; 21 AA.
XX
XX AC AAG62535;
XX
XX DT 24-AUG-2001 (first entry)
XX
XX DE Human CRH releasing protein related peptide SEQ ID NO: 36.
XX
XX KW Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
XX analgesic; hyperaldosteronism; hypercortisolism; hypoadrenocorticism;
XX Addison's disease; adrenal gland hyperfunction; obesity.
XX
XX OS Homo sapiens.
XX
XX PN WO200135984-A1.
XX
XX PD 25-MAY-2001.
XX
XX PF 17-NOV-2000; 2000WO-JP08119.
XX
XX PR 18-NOV-1999; 99JP-0327900.
XX
XX PR 26-SEP-2000; 2000JP-0297073.
XX

PA (TAKE) TAKEDA CHEM IND LTD.
XX
XX PI Kitada C, Matsumoto H, Hinuma S;
XX WPI; 2001-355552/37.
XX
XX DR
XX PT Use of G protein receptor ligand or peptide for controlling
XX corticotrophin releasing hormone secretion -
XX
XX PS Disclosure; Page 75; 90pp; Japanese.
XX
XX CC The present sequence describes a method of controlling the secretion of
XX corticotrophin releasing hormone (CRH), involving the use of a G protein
XX receptor ligand. This can be used to control the secretion of CRH and its
XX useful as an analgesic or for treating, preventing or ameliorating
XX diseases associated with CRH secretion such as hyperaldosteronism,
XX hypercortisolism, secondary or chronic hypoadrenocorticism, Addison's
XX disease (including, boredom, nausea, pigmentation, hypogonadism, hair
XX loss, and hypotension), adrenal gland hypofunction and obesity. The
XX present sequence is a peptide used in the exemplification of the
XX invention.
XX
XX SQ Sequence 21 AA;
XX
XX Query Match 96.3%; Score 105; DB 22; Length 21;
XX Best Local Similarity 90.0%; Pred. No. 3,4e-11;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 TPDIINAWYXXRGIRVGRF 20
XX | | | | | | | | | | | | | | | | | | | | | |
XX 1 TPDIINAWYASRGIRVGRF 20
XX
XX DB
XX
XX RESULT 10
XX AAW31396
XX ID AAW31396 standard; Peptide; 22 AA.
XX
XX AC AAW31396;
XX
XX DT 06-APR-1998 (first entry)
XX
XX DE Human type G protein-coupled receptor ligand fragment 6.
XX
XX KW G protein-coupled receptor; ligand binding; pharmaceutical;
XX modulator; pituitary; central nervous system; pancreas; prophylactic;
XX therapeutic agent.
XX
XX OS Homo sapiens.
XX
XX PN WO9724436-A2.
XX
XX PD 10-JUL-1997.
XX
XX PF 26-DEC-1996; 96WO-JP03821.
XX
XX PR 18-SEP-1996; 96JP-0246573.
XX
XX PR 28-DEC-1995; 95JP-0343371.
XX
XX PR 15-MAR-1996; 96JP-0059419.
XX
XX PR 12-AUG-1996; 96JP-0211805.
XX
XX PA (TAKE) TAKEDA CHEM IND LTD.
XX
XX PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
XX Kawamata Y, Kitada C;
XX WPI; 1997-363672/33.
XX
XX DR N-PSDB; AAW02433.
XX
XX PT Ligand peptide for G protein-coupled receptor - acts by modulating
XX function in the central nervous system, pancreas and pituitary gland
XX
XX PS Claim 2; Page 186; 256pp; English.
XX

CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 34 to 55 of the
 CC sequence represented in AAM31390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and hypophagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia, renal disease,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, spinal injury,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, splenocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharide. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.

CC Sequence 22 AA;

Query Match 96.3%; Score 105; DB 18; Length 22;
 Best Local Similarity 90.0%; Pred. No. 3.5e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXKRGIRPVGRF 20
 DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 11
 AAB10367
 ID AAB10367 standard; peptide; 22 AA.

AC AAB10367;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 37.

XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.

OS Homo sapiens.

XX WO2000038704-A1.

XX 06-JUL-2000,

XX 22-DEC-1999; 99WO-JP07199.

XX 25-DEC-1998; 98JP-0369585.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Matsumoto H, Kitada C, Hinuma S;

XX WPI; 2000-452298/39.

XX Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine

XX Disclosure; Page 64; 72pp; Japanese.

XX This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.

CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.

XX Sequence 22 AA;

Query Match 96.3%; Score 105; DB 21; Length 22;
 Best Local Similarity 90.0%; Pred. No. 3.5e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXKRGIRPVGRF 20
 DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 12
 AAG62536
 ID AAG62536 standard; peptide; 22 AA.

AC AAG62536;

DT 24-AUG-2001 (first entry)

DE Human CRH releasing protein related peptide SEQ ID NO: 37.

XX Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
 KW analgesic; hyperaldosteronism; hypercortisolaemia; hypoadrenocorticism;
 KW Addison's disease; adrenal gland hyperfunction; obesity.

OS Homo sapiens.

XX WO200135984-A1.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-JP08119.

XX 18-NOV-1999; 99JP-03227900.

XX 26-SEP-2000; 2000JP-0297073.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Kitada C, Matsumoto H, Hinuma S;

XX WPI; 2001-35552/37.

XX Use of G protein receptor ligand or peptide for controlling
 PT corticotrophin releasing hormone secretion -
 PT Disclosure; Page 75; 90pp; Japanese.

XX The present sequence describes a method of controlling the secretion of
 CC corticotrophin releasing hormone (CRH), involving the use of a G protein
 CC receptor ligand. This can be used to control the secretion of CRH and is
 CC useful as an analgesic or for treating, preventing or ameliorating
 CC diseases associated with CRH secretion such as hyperaldosteronism,
 CC hypercortisolaemia, secondary or chronic hypoadrenocorticism, Addison's
 CC disease (including boredom, nausea, pigmentation, hypogonadism, hair
 CC loss, and hypotension), adrenal gland hypofunction and obesity. The
 CC present sequence is a peptide used in the exemplification of the
 CC invention.

XX Sequence 22 AA;

Query Match 96.3%; Score 105; DB 22; Length 22;
 Best Local Similarity 90.0%; Pred. No. 3.5e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXKRGIRPVGRF 20
 DB 1 TPDINPAMYASRGIRPVGRF 20

DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 13

AAW31391 standard; Peptide: 31 AA.

AAW31391;

06-APR-1998 (first entry)

Human type G protein-coupled receptor ligand fragment 1.

G protein-coupled receptor; ligand binding; pharmaceutical; modulator; pituitary; central nervous system; pancreas; prophylactic; therapeutic agent.

Homo sapiens.

MO9724436-A2.

10-JUL-1997.

26-DEC-1996; 96WO-JP03821.

18-SEP-1996; 96JP-0246573.

28-DEC-1995; 95JP-0343371.

15-MAR-1996; 96JP-00594419.

12-AUG-1996; 96JP-0211805.

(TAKE) TAKEDA CHEM IND LTD.

Fuji R, Fukusumi S, Habata Y, Hinuma S, Hosoya M, Kawamata Y, Kitada C;

WPI: 1997-363672/33.

N-PSDB; AAV02428.

Claim 2; Page 184; 258pp; English.

This sequence represents a peptide fragment from a novel human type

ligand polypeptide corresponding to amino acid residues 23 to 53 of the

sequence represented in AAW31390 and is used in an assay to monitor

ligand binding to the G protein-coupled receptor protein. Pharmaceutical

compositions containing this ligand may be used as a pituitary function

modulator, a central nervous system modulator or a pancreatic function

modulator. This ligand could have specific applications as a

prophylactic or therapeutic agent for dementia, depression, hyperkinetic

syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,

trauma, growth hormone secretory disease, hyper- and polypnea,

hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,

hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,

Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,

transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,

acute myocardial infarction, infertility, spinocerebellar degeneration,

bone fracture, trauma, atopic dermatitis, osteoporosis and/or

oligodactyia. Assays can also be developed to screen compounds which are

capable of altering the binding activity of the ligand affecting

activation of the G protein-coupled receptor protein.

Sequence 31 AA;

Query Match

Best Local Similarity

Matches 18; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

RESULT 14

AAW97235 standard; peptide: 31 AA.

AAW97235;

06-MAY-1999 (first entry)

Human type ligand polypeptide fragment.

Rat type ligand; modulation; prolactin secretion;

G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacosgenesis;

menopausal syndrome; euthyroid; hypometabolism; lactation;

pituitary adenomatosis; brain tumour; amenorrhoea; galactorrhea;

prolactinoma; infertility; impotence; amenorrhoea; galactorrhea;

acromegaly; Chari-Frommel syndrome; Argon-del Castillo syndrome;

Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dysospermia;

contraceptive; placental function; chorioncarcinoma; hydatid mole;

irruption mole; abortion; unfertilized fetus; abnormal saccharometabolism;

abnormal lipidmetabolism; oxytocia.

Homo sapiens.

MO9858962-A1.

30-DEC-1998.

22-JUN-1998; 98WO-JP02765.

23-JUN-1997; 97JP-0165437.

(TAKE) TAKEDA CHEM IND LTD.

Fuji R, Hinuma S, Kawamata Y, Matsumoto H;

WPI: 1999-105614/09.

Claim 3; Page 159; 241pp; English.

The present sequence represents a human type ligand fragment. It

is used in the course of the invention. The specification describes

an agent for modulating prolactin secretion which comprises a

ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)

protein. The agents for promoting prolactin secretion can be used for

treating or preventing hypovarianism, gonocyst cacosgenesis, menopausal

syndrome, euthyroid or hypometabolism. They can be used for promoting

lactation in a domestic mammal and as an aphrodisiac. The agents for

inhibiting prolactin secretion can be used for treating or preventing

pituitary adenomatosis, brain tumour, emmenorrhoea, autoimmune disease,

prolactinoma, infertility, impotence, amenorrhoea, galactorrhea,

acromegaly, Chari-Frommel syndrome, Argon-del Castillo syndrome,

Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dysospermia.

The inhibitory agents can also be used as contraceptives. The agents for

modulating placental function can be used for treating or preventing

choriocarcinoma, hydatid mole, irruption mole, abortion, unfertilized fetus,

abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

Sequence 31 AA;

Query Match

Best Local Similarity

Matches 18; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

RESULT 15

AAW87615 standard; Peptide; 31 AA.

AC AAW87615;

DT 29-MAR-1999 (first entry)

DE Human 19P2 ligand.

KW 19P2 ligand; G protein coupled receptor; pituitary; prolactin releasing peptide; human; dementia; breast cancer; therapy.

OS Homo sapiens.

PN EP887417-A2.

PD 30-DEC-1998.

PF 25-JUN-1998; 98EP-0111725.

PR 27-JUN-1997; 97JP-0172118.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;

DR WPI, 1999-047884/05.

PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage of a fusion protein, useful for preventing and treating dementia, breast cancer, renal failure and autoimmune disease

PS Claim 5; Page 35; 56pp; English.

XX This is the amino acid sequence of the human pituitary G protein-coupled receptor ligand 19P2L. A method suitable for commercial high-level production of 19P2L comprises expressing the ligand in host cells as a recombinant fusion protein e.g. with human basic fibroblast growth factor (see AY83796-97) that has been modified to include an N-terminal cysteine residue. The ligand is released from the fusion by cyanidation followed by ammonolysis. 19P2L has prolactin secretion-stimulating and (at high doses) prolactin secretion-inhibiting properties. It can be used in the treatment and prevention of various diseases including: senile dementia, cerebrovascular dementia, and dementia associated with: neurological disorders (e.g. Alzheimer's disease, Parkinson's disease, Pick's disease, Huntington's disease), infectious diseases (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism, intoxication by drugs, metal and organic compounds), tumourigenic diseases (e.g. brain tumour), traumatic diseases (e.g. chronic subarachnoid haemorrhage, and other types of dementia, depression, hyperactive child syndrome (microcephalopathy) and disturbance of consciousness. It is also useful for prevention and treatment of diseases associated with prolactin hypo and hypersecretion respectively, including: hyperprolactinaemia, pituitary adenoma, breast cancer, infertility, impotence and autoimmune disease (hypersecretion disorders), and seminal vesicle hypoplasia, osteoporosis, menopausal syndrome and renal failure (hyposecretion disorders). The 19P2 polypeptide/amide is also useful as a test reagent for study of the prolactin secretory function or as a lactagogue in mammalian farm animals.

SQ Sequence 31 AA;

Query Match 96.3%, score 105, DB 20, length 31,

Best Local Similarity 90.0%, Pred. No. 5.2e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPANTYASRGIRPVGRF 20

|||||

DB 12 TPDINPANTYASRGIRPVGRF 31

Search completed: May 1, 2003, 14:30:50
Job time : 31.5938 secs

Thu May 1 15:54:27 2003

Re-run

US-09-446-543a-74.1rag

SEQ ID NO: 74
Database: A-Geneseg-101002
AC NO: AAW31384

File copy
Page 1
Cited in
paper 4-16

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 15.2969 Seconds
(without alignments)
95.821 Million cell updates/sec

Title: US-09-446-543a-74
Perfect score: 47
Sequence: 1 SRXHSMEKX 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A-Geneseg-101002:

- 1: /SID2/gcgdata/geneseg/genesegp-emb1/AA1980.DAT*
- 2: /SID2/gcgdata/geneseg/genesegp-emb1/AA1981.DAT*
- 3: /SID2/gcgdata/geneseg/genesegp-emb1/AA1982.DAT*
- 4: /SID2/gcgdata/geneseg/genesegp-emb1/AA1983.DAT*
- 5: /SID2/gcgdata/geneseg/genesegp-emb1/AA1984.DAT*
- 6: /SID2/gcgdata/geneseg/genesegp-emb1/AA1985.DAT*
- 7: /SID2/gcgdata/geneseg/genesegp-emb1/AA1986.DAT*
- 8: /SID2/gcgdata/geneseg/genesegp-emb1/AA1987.DAT*
- 9: /SID2/gcgdata/geneseg/genesegp-emb1/AA1988.DAT*
- 10: /SID2/gcgdata/geneseg/genesegp-emb1/AA1989.DAT*
- 11: /SID2/gcgdata/geneseg/genesegp-emb1/AA1990.DAT*
- 12: /SID2/gcgdata/geneseg/genesegp-emb1/AA1991.DAT*
- 13: /SID2/gcgdata/geneseg/genesegp-emb1/AA1992.DAT*
- 14: /SID2/gcgdata/geneseg/genesegp-emb1/AA1993.DAT*
- 15: /SID2/gcgdata/geneseg/genesegp-emb1/AA1994.DAT*
- 16: /SID2/gcgdata/geneseg/genesegp-emb1/AA1995.DAT*
- 17: /SID2/gcgdata/geneseg/genesegp-emb1/AA1996.DAT*
- 18: /SID2/gcgdata/geneseg/genesegp-emb1/AA1997.DAT*
- 19: /SID2/gcgdata/geneseg/genesegp-emb1/AA1998.DAT*
- 20: /SID2/gcgdata/geneseg/genesegp-emb1/AA1999.DAT*
- 21: /SID2/gcgdata/geneseg/genesegp-emb1/AA2000.DAT*
- 22: /SID2/gcgdata/geneseg/genesegp-emb1/AA2001.DAT*
- 23: /SID2/gcgdata/geneseg/genesegp-emb1/AA2002.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	91.5	31	18	AAW31384
2	43	91.5	31	20	AAW97233
3	43	91.5	31	20	AAW87614
4	43	91.5	31	20	AAW95173
5	43	91.5	31	20	AAW95174
6	43	91.5	31	21	AAW10355
7	43	91.5	31	21	AAW87504
8	43	91.5	31	21	AAW49292
9	43	91.5	31	22	AAW62524
10	43	91.5	31	22	AAW90993

11	43	91.5	32	18	AAW31385	Rat type G protein
12	43	91.5	32	21	AAW10356	Rat oxytocin secre
13	43	91.5	32	22	AAW62525	Rat CRH releasing
14	43	91.5	33	18	AAW31386	Rat type G protein
15	43	91.5	33	21	AAW10357	Rat oxytocin secre
16	43	91.5	33	22	AAW62526	Rat CRH releasing
17	43	91.5	32	18	AAW95172	Murine pituitary-d
18	43	91.5	33	20	AAW31383	Rat type G protein
19	43	91.5	33	21	AAW97235	Rat oxytocin secre
20	43	91.5	33	22	AAW10354	Rat CRH releasing
21	43	91.5	33	21	AAW10355	Synthetic ligand 1
22	42	89.4	15	20	AAW31399	Bovine G protein-c
23	42	89.4	15	18	AAW97229	Bovine pituitary-d
24	42	89.4	29	18	AAW31369	Human type G prote
25	42	89.4	29	20	AAW95184	Bovine G protein-c
26	42	89.4	30	21	AAW49299	Human type G prote
27	42	89.4	31	18	AAW31391	Bovine G protein-c
28	42	89.4	31	20	AAW97235	Human type ligand
29	42	89.4	31	20	AAW97218	Bovine pituitary-d
30	42	89.4	31	20	AAW87613	Bovine 1992 ligand
31	42	89.4	31	20	AAW87615	Human 1992 ligand
32	42	89.4	31	20	AAW95188	Bovine pituitary-d
33	42	89.4	31	21	AAW10347	Bovine oxytocin se
34	42	89.4	31	21	AAW10362	Bovine oxytocin se
35	42	89.4	31	21	AAW49290	1992 ligand peptid
36	42	89.4	31	21	AAW49291	1992 ligand peptid
37	42	89.4	31	21	AAW49298	Bovine CRH releasi
38	42	89.4	31	21	AAW62531	Human CRH releasi
39	42	89.4	31	22	AAW62516	Proactin releasin
40	42	89.4	31	22	AAW90991	Proactin releasin
41	42	89.4	31	22	AAW90995	bPrp31 peptide, u
42	42	89.4	31	22	AAW31370	Human type G prote
43	42	89.4	31	22	AAW31392	Bovine G protein-c
44	42	89.4	32	18	AAW31372	
45	42	89.4	32	18	AAW31372	

ALIGNMENTS

RESULT 1	AAW31384	standard; Peptide; 31 AA.
ID	AAW31384;	
AC	AAW31384;	
XX		
DT	06-APR-1998 (first entry)	
XX		
DE	Rat type G protein-coupled receptor ligand fragment 1.	
XX		
KW	G protein-coupled receptor; ligand binding; Pharmacological;	
KM	modulator; pituitary; central nervous system; pancreas; prophylactic;	
KW	therapeutic agent.	
XX		
OS	Rat sp.	
XX		
PN	W09724436-A2	
XX		
PD	10-JUL-1997.	
XX		
PF	26-DEC-1996; 96W0-JP03821.	
XX		
PR	18-SEP-1996; 96GP-0246573.	
PR	28-DEC-1995; 95GP-0343371.	
PR	15-MAR-1996; 96GP-0059419.	
PR	12-AUG-1996; 96GP-0211805.	
XX		
PA	(TAKE) TAKEDA CHEM IND LTD.	
XX		
PI	Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M,	
PI	Kawamata Y, Kitada C;	
XX		
DR	WPI: 1997-363672/33.	

DR N-PSDB; AAW02421.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 PS Claim 2; Page 179; 258bp; English.

XX This sequence represents a peptide fragment from a novel rat type
 CC ligand polypeptide corresponding to amino acid residues 22 to 52 of the
 CC sequence represented in AAW31383 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC propylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreaticitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosacchara. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.

SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 18; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11

DB 1 SRAHQSMEXR 11

RESULT 2
 AAW97233
 ID AAW97233 standard; peptide; 31 AA.

AC AAW97233;

DT 06-MAY-1999 (first entry)

DE Rat type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
 KW G protein-coupled receptor; GPCR; hypocoarctanism; gonocyst coccogenesis;
 KW menopausal syndrome; euthyroid; hypometabolism; lactation;
 KW pituitary adenomatosis; brain tumour; amenorrhoea; galactorrhoea;
 KW prolactinoma; infertility; impotence; amenorrhoea; galactorrhoea;
 KW acromegaly; Chlari-Frommel syndrome; Argon2-del Castillo syndrome;
 KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
 KW contraceptive; placental function; choriochorionoma; hydatid mole;
 KW abortion; unfertilized fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.

XX Rattus sp.

XX W09858962-A1.

XX 30-DEC-1998.

XX 22-JUN-1998; 98WO-JP02765.

XX 23-JUN-1997; 97JP-0155437.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Fujii R, Hinuma S, Kawamata Y, Matsumoto N;

DR WPI; 1999-105614/09.

XX Use of G protein-coupled receptor ligands - for modulating prolactin
 PT secretion or placental function, e.g. for treating menopausal
 PT syndrome, tumours, autoimmune disease or abnormal pregnancy
 PS Claim 3; Page 153; 241bp; English.

XX The present sequence represents a rat type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypocoarctanism, gonocyst coccogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, amenorrhoea, galactorrhoea,
 CC prolactinoma, infertility, impotence, amenorrhoea, galactorrhoea,
 CC acromegaly, Chlari-Frommel syndrome, Argon2-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC choriochorionoma, hydatid mole, abortion, unfertilized fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 20; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11

DB 1 SRAHQSMEXR 11

RESULT 3
 AAW87614
 ID AAW87614 standard; peptide; 31 AA.

AC AAW87614;

DT 29-MAR-1999 (first entry)

DE Rat 19P2 ligand.

XX 19P2 ligand; G protein coupled receptor; pituitary;
 KW prolactin releasing peptide; rat; dementia; breast cancer;
 KW therapy.

XX Rattus sp.

XX EP887417-A2.

XX 30-DEC-1998.

XX 25-JUN-1998; 98EP-0111725.

XX 27-JUN-1997; 97JP-0172118.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Moriya T, Nishimura O, Suenaga M, Tanaka Y;

XX WPI; 1999-047884/05.

XX Producing a 19P2 pituitary G protein receptor ligand - by cleavage
 PT of a fusion protein, useful for preventing and treating dementia,
 PT breast cancer, renal failure and autoimmune disease
 PS Claim 5; Page 34; 56bp; English.

CC This is the amino acid sequence of the rat pituitary G
 CC protein-coupled receptor ligand 19p2L. A method suitable for
 CC commercial high-level production of 19p2L comprises expressing
 CC the ligand in host cells as a recombinant fusion protein e.g. with
 CC human basic fibroblast growth factor (see A463794-95) that has
 CC been modified to include an N-terminal cysteine residue. The
 CC ligand is released from the fusion by cyanylation followed by
 CC ammonolysis. 19p2L has prolactin secretion-stimulating and (at
 CC high doses) prolactin secretion-inhibiting properties. It can be
 CC used in the treatment and prevention of various diseases including:
 CC senile dementia, cerebrovascular dementia, and dementia associated
 CC with: neurological disorders (e.g. Alzheimer's disease, Parkinson's
 CC disease, Pick's disease, Huntington's disease), infectious diseases
 CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
 CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
 CC intoxication by drugs, metal and organic compounds), tumorigenic
 CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
 CC subarachnoidal haemorrhage, and other types of dementia, depression,
 CC hyperactive child syndrome (microencephalopathy) and disturbance of
 CC consciousness. It is also useful for prevention and treatment of
 CC diseases associated with prolactin hypo and hypersecretion
 CC respectively, including: hyperprolactinaemia, pituitary adenoma,
 CC breast cancer, infertility, impotence and autoimmune disease
 CC (hypersecretion disorders), and seminal vesicle hypoplasia.
 CC osteoporosis, menopausal syndrome and renal failure (hypersecretion
 CC disorders). The 19p2 polypeptide/amide is also useful as a test
 CC reagent for study of the prolactin secretory function or as a
 CC lactagogue in mammalian farm animals.

CC Sequence 31 AA;

Query Match 91.5%; Score 43; DB 20; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SRXHSMEKRR 11
 || || || || ||
 Db 1 SRXHSMEKRR 11

RESULT 4
 AAW95173
 ID AAW95173 standard; peptide; 31 AA.

AC AAW95173;

DT 10-MAR-1999 (first entry)

DE Murine pituitary-derived ligand mature polypeptide sequence.

KM Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
 KM GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
 KM tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
 KM Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
 KM Creutzfeldt-Jakob disease; poisoning; schizophrenia; growth hormone;
 KM secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
 KM gene therapy; transgenic animal.

OS Mus sp.

PN W09849295-A1.

PD 05-NOV-1998.

PF 27-APR-1998; 98WO-JP01923.

PR 28-APR-1997; 97JP-0109974.

XX (TAKE) TAKEDA CHEM IND LTD.

PA Fukusumi S, Hinuma S;

PI WPI; 1999-009423/01.

XX New polypeptide ligand for orphan G protein coupled receptors - used
 PT for treating disorders of central nervous system, pituitary and
 PT pancreas, and for drug screening

PS Disclosure; Page 134; 206pp; English.

XX This represents the matured murine pituitary-derived ligand polypeptide
 CC sequence. The polypeptide is a ligand for the G-protein coupled orphan
 CC receptor designated GPR10 (human) or UHR-1 (rat). Cells transformed with
 CC a vector containing the ligand polypeptide encoding DNA are used to
 CC produce a recombinant ligand polypeptide. The ligand polypeptide, and its
 CC fragments, modulate function of the pituitary, central nervous system,
 CC pancreas and other tissues and can be used to screen for agents that
 CC modulate binding of the polypeptide to the receptor; to quantify the
 CC amount of receptor in a sample and to raise antibodies. They may also be
 CC used therapeutically, e.g. to treat senile dementia; Alzheimer's,
 CC Parkinson's or Huntington's diseases; Creutzfeldt-Jakob disease; poisoning
 CC by heavy metals or drugs; diabetes; schizophrenia; disorders of growth
 CC hormone secretion; cancer; rheumatoid arthritis, epilepsy and many
 CC others, also to improve post-operative nutritional status and as
 CC vasopressor. Transgenic animals carrying the ligand polypeptide encoding
 CC DNA or its mutant are used to study the function of the polypeptide-
 CC expressing genes, as models of disease, for drug screening and as source
 CC of cell lines. The ligand polypeptide DNA is used as a source of probes
 CC and primers; to identify related sequences; in receptor-binding assays;
 CC for production of Ab and antisera; in drug development; for gene therapy
 CC and to develop transgenic animals.

CC Sequence 31 AA;

Query Match 91.5%; Score 43; DB 20; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SRXHSMEKRR 11
 || || || || ||
 Db 1 SRXHSMEKRR 11

RESULT 5
 AAW95174
 ID AAW95174 standard; Protein; 31 AA.

AC AAW95174;

DT 10-MAR-1999 (first entry)

DE Murine pituitary-derived ligand polypeptide antigenic epitope.

KM Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
 KM GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
 KM tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
 KM Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
 KM Creutzfeldt-Jakob disease; poisoning; schizophrenia; growth hormone;
 KM secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
 KM gene therapy; transgenic animal; epitope.

OS Mus sp.

PN W09849295-A1.

PD 05-NOV-1998.

PF 27-APR-1998; 98WO-JP01923.

PR 28-APR-1997; 97JP-0109974.

XX (TAKE) TAKEDA CHEM IND LTD.

PA Fukusumi S, Hinuma S;

PI WPI; 1999-009423/01.

CC the activity of GPR10, a G protein-coupled receptor. These assays are
 CC then followed by an in vivo assay of the effect of the compound on
 CC feeding behaviour, body weight or metabolic rate in a mammal. Prolactin-
 CC releasing peptide (PrRP; AAY97504) is a ligand of GPR10. Binding of PrRP
 CC to GPR10 stimulates a signal transduction cascade, which results in an
 CC increase in food intake. Compounds identified using the method of the
 CC invention are useful for the modulation of body weight. Antagonists of
 CC GPR10 can be used to treat obesity, while GPR10 agonists can be used to
 CC treat cachexia. The present sequence represents rat PrRP.

XX Sequence 31 AA;

Query Match 91.5%; Score 43; DB 21; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRKHSHMEKR 11
 DB 1 SRKHSHMEKR 11

RESULT 8
 AAY49292

ID AAY49292 standard; peptide; 31 AA.

XX AAY49292;

DT 22-FEB-2000 (first entry)

DE 19P2 ligand peptide fragment.

KW Monoclonal antibody, 19P2 ligand; diagnosis; prolactin secretion;

KW pituitary; regulatory mechanism; central nervous system; pancreatic.

XX Rattus sp.

XX Key Location/Qualifiers

FT Modified-site 31 /note- "C-terminal amide"

PN W09960112-A1.

XX 25-NOV-1999.

PP 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

XX (TAKE) TAKEDA CHEM IND LTD.

PI Katsumoto H, Kitada C, Hinuma S;

XX WPI; 2000-039361/03.

PT New monoclonal antibodies, useful in diagnosis, as drugs and in
 PT studying diseases related to ligand abnormality

XX Disclosure; Page 26; 73pp; Japanese.

CC The invention provides a monoclonal antibody which has a specific
 CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
 CC derivative. The antibodies can be used in diagnosis or to treat or
 CC prevent diseases associated with abnormality in the pituitary function
 CC regulatory mechanism (e.g. promotion of prolactin secretion), central
 CC nervous regulatory mechanism, and pancreatic function regulation
 CC mechanism. The antibody-based immunoassay can also be applied in
 CC clarifying the physiological functions of the ligand and its derivative.
 CC Sequences AAY49290-302 represent peptide fragments of the 19P2 ligand.

XX Sequence 31 AA;

Query Match 91.5%; Score 43; DB 21; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRKHSHMEKR 11
 DB 1 SRKHSHMEKR 11

RESULT 9
 AAG62524

ID AAG62524 standard; peptide; 31 AA.

XX AAG62524;

DT 24-AUG-2001 (first entry)

DE Rat CRH releasing protein related peptide SEQ ID NO: 18.

KW Rat; corticotrophin releasing hormone; CRH; G protein receptor ligand;

KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;

KW Addison's disease; adrenal gland hyperfunction; obesity.

XX Rattus sp.

XX W0200135984-A1.

XX 25-MAY-2001.

PE 17-NOV-2000; 2000WO-JP08119.

PR 18-NOV-1999; 99JP-0327900.

PR 26-SEP-2000; 2000JP-0297073.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Kitada C, Katsumoto H, Hinuma S;

XX WPI; 2001-355552/37.

PT Use of G protein receptor ligand or peptide for controlling

PT corticotrophin releasing hormone secretion

XX Claim 3; Page 69; 90pp; Japanese.

CC The present sequence describes a method of controlling the secretion of
 CC corticotrophin releasing hormone (CRH), involving the use of a G protein
 CC receptor ligand. This can be used to control the secretion of CRH and is
 CC useful as an analgesic or for treating, preventing or ameliorating
 CC diseases associated with CRH secretion such as hyperaldosteronism,
 CC hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
 CC disease (including borden, nausea, pigmentation, hypogonadism, hair
 CC loss, and hypotension), adrenal gland hyperfunction and obesity. The
 CC present sequence is a peptide used in the exemplification of the
 CC invention.

XX Sequence 31 AA;

Query Match 91.5%; Score 43; DB 22; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRKHSHMEKR 11
 DB 1 SRKHSHMEKR 11

RESULT 10
 AAB90993

ID AAB90993 standard; peptide; 31 AA.

XX AAB90993;

DT 22-JUN-2001 (first entry)

DE Prolactin releasing peptide SEQ ID NO:167.
 XX
 XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl, maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX
 OS Homo sapiens.
 OS Synthetic.
 PN WO200069900-A2.
 XX
 PD 23-NOV-2000.
 XX
 PE 17-MAY-2000; 2000WO-US13576.
 XX
 PR 17-MAY-1999; 9905-0134406.
 PR 10-SEP-1999; 9905-0153406.
 PR 15-OCT-1999; 9905-0159783.
 XX
 PA (CONF-) CONJUCHEM INC.
 XX
 PI Bridon DP, Ezrin AM, Malner PG, Holmes DL, Thibaudau K;
 DR WPI; 2001-112059/12.
 XX
 PT Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity
 PS Disclosure; Page 244; 733pp; English.
 XX
 CC The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity
 CC in vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAPP90829 to AAPP92441 represent peptides which can be used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 31 AA;
 XX
 QY Query Match 91.5%; Score 43; DB 22; Length 31;
 Db Best Local Similarity 72.7%; Pred. No. 0.031;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 SRXHXSMEXR 11
 II I I I I I I
 Db 1 SRAHQSMETR 11
 XX
 RESULT 11
 AAB10356
 ID AAB10356 standard; peptide; 32 AA.
 AC AAB10356;
 XX
 DT 06-APR-1998 (first entry)
 XX
 OE Rat type G protein-coupled receptor ligand fragment 2.
 XX
 KW G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.

XX
 OS Rat sp.
 XX
 PN WO9724436-A2.
 XX
 PD 10-JUL-1997.
 XX
 PE 26-DEC-1996; 96WO-JP03821.
 XX
 PR 18-SEP-1996; 96JP-0246573.
 PR 28-DEC-1995; 95JP-0343371.
 PR 15-MAR-1996; 95JP-0059419.
 PR 12-AUG-1996; 96JP-0211805.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
 PI Kawamata Y, Kikada G;
 DR N-PSDB; AAV02422.
 XX
 PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 PS Claim 2; Page 179; 258pp; English.
 XX
 CC This sequence represents a peptide fragment from a novel rat type
 CC ligand polypeptide corresponding to amino acid residues 22 to 53 of the
 CC sequence represented in AAB10356 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neuritis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharide. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 XX
 SQ Sequence 32 AA;
 XX
 QY Query Match 91.5%; Score 43; DB 18; Length 32;
 Db Best Local Similarity 72.7%; Pred. No. 0.032;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 SRXHXSMEXR 11
 II I I I I I I
 Db 1 SRAHQSMETR 11
 XX
 RESULT 12
 AAB10356
 ID AAB10356 standard; peptide; 32 AA.
 AC AAB10356;
 XX
 DT 24-NOV-2000 (first entry)
 XX
 OE Rat oxytocin secretion promoting peptide SEQ ID NO: 19.
 XX
 KW Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.

OS Rattus sp.
 PN WO200038704-A1.
 XX 06-JUL-2000.
 XX
 PE 22-DEC-1999; 99WO-JP07199.
 XX
 PR 25-DEC-1998; 98JP-0369585.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Matsumoto H, Kitada C, Hinuma S;
 DR WPI; 2000-452298/39.
 XX
 PT Physiologically active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine -
 XX
 PS Disclosure; Page 57; 72pp; Japanese.
 XX
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.,
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a rat peptide which acts as an oxytocin secretion
 CC promoter.
 XX
 SQ Sequence 32 AA;
 XX
 Query Match 91.5%; Score 43; DB 21; Length 32;
 Best Local Similarity 72.7%; Pred. No. 0.032;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 || || || || ||
 Db 1 SRAHSHSMETR 11
 XX
 RESULT 13
 ID AAG62525
 XX AAG62525 standard; peptide; 32 AA.
 XX
 AC AAG62525;
 XX
 DT 24-AUG-2001 (first entry)
 XX
 DE Rat CRH releasing protein related peptide SEQ ID NO: 19.
 XX
 DE Rat; corticotrophin releasing hormone; CRH; G protein receptor ligand;
 KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
 KW Addison's disease; adrenal gland hyperfunction; obesity.
 XX
 OS Rattus sp.
 XX
 PN WO200135984-A1.
 XX
 PD 25-MAY-2001.
 XX
 PE 17-NOV-2000; 2000WO-JP08119.
 XX
 PR 18-NOV-1999; 99JP-0327900.
 XX
 PR 26-SEP-2000; 2000JP-0297073.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Kitada C, Matsumoto H, Hinuma S;

XX
 DR WPI; 2001-355552/37.
 XX
 PT Use of G protein receptor ligand or peptide for controlling
 PT corticotrophin releasing hormone secretion -
 XX
 PS Disclosure; Page 69; 90pp; Japanese.
 XX
 CC The present sequence describes a method of controlling the secretion of
 CC corticotrophin releasing hormone (CRH), involving the use of a G protein
 CC receptor ligand. This can be used to control the secretion of CRH and is
 CC useful as an analgesic or for treating, preventing or ameliorating
 CC diseases associated with CRH secretion such as hyperaldosteronism,
 CC hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
 CC disease (including boredom, nausea, pigmentation, hypogonadism, hair
 CC loss, and hypotension), adrenal gland hypofunction and obesity. The
 CC present sequence is a peptide used in the exemplification of the
 CC invention.
 XX
 SQ Sequence 32 AA;
 XX
 Query Match 91.5%; Score 43; DB 22; Length 32;
 Best Local Similarity 72.7%; Pred. No. 0.032;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 || || || || ||
 Db 1 SRAHSHSMETR 11
 XX
 RESULT 14
 ID AAW31386
 XX AAW31386 standard; Peptide; 33 AA.
 XX
 AC AAW31386;
 XX
 DT 06-APR-1998 (first entry)
 XX
 DE Rat type G protein-coupled receptor ligand fragment 3.
 XX
 DE G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.
 XX
 OS Rat sp.
 XX
 PN WO9724436-A2.
 XX
 PD 10-JUL-1997.
 XX
 PE 26-DEC-1996; 96WO-JP03821.
 XX
 PR 18-SEP-1996; 96JP-0246573.
 XX
 PR 28-DEC-1995; 95JP-0343371.
 XX
 PR 15-MAR-1996; 96JP-0059419.
 XX
 PR 12-AUG-1996; 96JP-0211805.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
 PI Kawamata Y, Kitada C;
 DR WPI; 1997-363672/33.
 DR H-PSDB; AAW02423.
 XX
 PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2; Page 179-180; 258pp: English.
 XX
 CC This sequence represents a peptide fragment from a novel rat type
 CC ligand polypeptide corresponding to amino acid residues 22 to 54 of the
 CC sequence represented in AAW31383 and is used in an assay to monitor

CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophyllactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharide. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 CC
 XX Sequence 33 AA:

Query Match 91.5%; Score 43; DB 18; Length 33;
 Best Local Similarity 72.7%; Pred. No. 0.033;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSHMEXR 11
 || || || || ||
 Db 1 SRAHQSHMETR 11

RESULT 15
 AAB10357

ID AAB10357 standard; peptide; 33 AA.

XX AAB10357;

DT 24-NOV-2000 (first entry)

DE Rat oxytocin secretion promoting peptide SEQ ID NO: 20.

XX Rat; oxytocin secretion promoter; G protein-coupled receptor protein;

KM treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;

KM caesarean section; artificial fertilization; galactostasis; goat; pig;

XX veterinary medicine; milk production.

OS Rattus sp.

PN W0200038704-A1.

PD 06-JUL-2000.

XX 22-DEC-1999; 99MO-JP07199.

XX 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

XX Physiologically active polypeptide recognized as ligand by G

PT protein-coupled receptor protein, for promoting secretion of oxytocin,

PT as drugs for diseases relating to oxytocin secretion and in veterinary

PT medicine

XX Disclosure; Page 58; 72pp; Japanese.

XX This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary

CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a rat peptide which acts as an oxytocin secretion
 CC promoter.

XX Sequence 33 AA:

Query Match 91.5%; Score 43; DB 21; Length 33;
 Best Local Similarity 72.7%; Pred. No. 0.033;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSHMEXR 11
 || || || || ||
 Db 1 SRAHQSHMETR 11

Search completed: May 1, 2003, 14:30:50
 Job time : 15.2969 secs

File Copy
Thu May 1 15:54:02 2003
Cited in paper #16
de-lun

GenCore version 5.1.4-P5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:36:27 ; Search time 35 Seconds
(without alignments)
125.636 Million cell updates/sec

Title: SE074-PLUS-73
Sequence: 1 SRKXHSMEKRPDIPNPNVYKXKIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.0

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneq/geneq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneq/geneq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneq/geneq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneq/geneq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneq/geneq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneq/geneq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneq/geneq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneq/geneq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneq/geneq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneq/geneq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneq/geneq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneq/geneq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneq/geneq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneq/geneq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneq/geneq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneq/geneq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneq/geneq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneq/geneq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneq/geneq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneq/geneq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneq/geneq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneq/geneq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneq/geneq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	31	18	AAW31391
2	147	94.2	31	18	AAW31384
3	147	94.2	31	20	AAW97233
4	147	94.2	31	20	AAW97235
5	147	94.2	31	20	AAW87614
6	147	94.2	31	20	AAW87615
7	147	94.2	31	20	AAW95173
8	147	94.2	31	20	AAW95174
9	147	94.2	31	21	AAW10355
10	147	94.2	31	21	AAW10362

SEQ ID NO: 74 fused to SEQ ID NO: 73
Database: A-GeneSeq-101002
AC NO: AAW31391
Page 1

11	147	94.2	31	21	AAW87504	Rat prolactin-rele
12	147	94.2	31	21	AAW43291	19P2 ligand peptid
13	147	94.2	31	21	AAW43292	19P2 ligand peptid
14	147	94.2	31	22	AAW62524	Rat CRH releasing
15	147	94.2	31	22	AAW62531	Human CRH releasing
16	147	94.2	31	22	AAW90991	prolactin releasin
17	147	94.2	31	22	AAW90993	prolactin releasin
18	147	94.2	31	22	AAW90995	prolactin releasin
19	147	94.2	31	22	AAW31392	Human type G prote
20	147	94.2	31	18	AAW31385	Human type G prote
21	147	94.2	32	21	AAW10356	Rat oxytocin secre
22	147	94.2	32	21	AAW10363	Rat oxytocin secre
23	147	94.2	32	22	AAW62525	Human oxytocin sec
24	147	94.2	32	22	AAW62532	Rat CRH releasing
25	147	94.2	33	18	AAW31393	Human CRH releasin
26	147	94.2	33	18	AAW31386	Human type G prote
27	147	94.2	33	21	AAW10357	Rat type G protein
28	147	94.2	33	21	AAW10364	Rat oxytocin secre
29	147	94.2	33	22	AAW62526	Human oxytocin sec
30	147	94.2	33	22	AAW62533	Rat CRH releasing
31	147	94.2	82	20	AAW95172	Human CRH releasin
32	147	94.2	83	18	AAW31383	Murine pituitary-d
33	147	94.2	83	20	AAW97225	Rat type G protein
34	147	94.2	83	21	AAW10354	Rat type ligand po
35	147	94.2	83	21	AAW62523	Rat oxytocin secre
36	147	94.2	87	18	AAW31390	Human type G prote
37	147	94.2	87	20	AAW97226	Human type ligand
38	147	94.2	87	21	AAW10361	Human CRH releasin
39	147	94.2	87	22	AAW62530	Human oxytocin sec
40	146	93.6	31	18	AAW31371	Bovine G protein-c
41	146	93.6	31	20	AAW97218	Bovine pituitary-d
42	146	93.6	31	20	AAW87613	Bovine 19P2 ligand
43	146	93.6	31	20	AAW95188	Bovine pituitary-d
44	146	93.6	31	21	AAW10347	Bovine oxytocin se
45	146	93.6	31	21	AAW49290	19P2 ligand peptid

ALIGNMENTS

RESULT 1	AAW31391	standard; Peptide: 31 AA.
ID	AAW31391	
AC	AAW31391	
XX		
DT	06-APR-1998	(first entry)
XX		
DE	Human type G protein-coupled receptor ligand fragment 1.	
XX		
RW	G protein-coupled receptor; ligand binding; pharmaceutical;	
RW	modulator; pituitary; central nervous system; pancreas; prophylactic;	
RW	therapeutic agent.	
XX		
OS	Homo sapiens.	
XX		
PN	W09724436-A2.	
XX		
PD	10-JUL-1997.	
XX		
FE	26-DEC-1996;	96NC-JP03821.
XX		
PR	18-SEP-1996;	96JP-0246573.
PR	28-DEC-1995;	95JP-0343371.
PR	15-MAR-1996;	96JP-0059419.
XX	12-AUG-1996;	96JP-0211805.
XX		
PA	(TAKE) TAKEDA CHEM IND LTD.	
XX		
PI	Fujita R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;	
PI	Kawamata Y, Kitada C;	
XX		
DR	RPI; 1997-363672/33.	

DR N-PSDB; AAV02428.
 XX Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2: Page 184; 258pp; English.
 XX
 CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 23 to 53 of the
 CC sequence represented in AAW31383 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator. This ligand could have specific applications as a
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC propylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 XX
 SQ Sequence 31 AA;
 XX
 OY Query Match 94.2%; Score 147; DB 18; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 XX
 DB 1 SRXHXSMKRTDPDINPANTYXXRGIRPVGRF 31
 1 SRTHRSMKRTDPDINPANTYASRGIRPVGRF 31
 XX
 RESULT 2
 AAW31384
 ID AAW31384 standard; peptide; 31 AA.
 AC AAW31384;
 XX
 DT 06-APR-1998 (first entry)
 XX
 DE Rat type G protein-coupled receptor ligand fragment 1.
 XX
 KW G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; propylactic;
 KW therapeutic agent.
 XX
 OS Rat sp.
 XX
 PN W09724436-A2.
 XX
 PD 10-JUL-1997.
 XX
 PE 26-DEC-1996; 96MO-JP03821.
 XX
 PR 18-SEP-1996; 96JP-0246573.
 PR 28-DEC-1995; 95JP-034371.
 PR 15-MAR-1996; 96JP-0059419.
 PR 12-AUG-1996; 96JP-0211805.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M,
 PI Kawamata Y, Kitada C;
 XX
 DR WPI; 1997-363672/33.
 DR N-PSDB; AAV02421.
 XX

PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2: Page 179; 258pp; English.
 XX
 CC This sequence represents a peptide fragment from a novel rat type
 CC ligand polypeptide corresponding to amino acid residues 22 to 52 of the
 CC sequence represented in AAW97233 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator. This ligand could have specific applications as a
 CC propylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 XX
 SQ Sequence 31 AA;
 XX
 OY Query Match 94.2%; Score 147; DB 18; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 XX
 DB 1 SRXHXSMKRTDPDINPANTYXXRGIRPVGRF 31
 1 SRAHOSMKRTDPDINPANTYGTGIRPVGRF 31
 XX
 RESULT 3
 AAW97233
 ID AAW97233 standard; peptide; 31 AA.
 AC AAW97233;
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE Rat type ligand polypeptide fragment.
 XX
 KW Rat type ligand; modulation; prolactin secretion;
 KW G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacoogenesis;
 KW menopausal syndrome; euthyroid; hypometabolism; lactation;
 KW pituitary adenomatosis; brain tumour; emmenopathy; autoimmune disease;
 KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
 KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
 KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
 KW contraceptive; placental function; chorioncarcinoma; hydralid mole;
 KW interruption mole; abortion; unfertilized fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.
 XX
 OS Rattus sp.
 XX
 PN W09858962-A1.
 XX
 PD 30-DEC-1998.
 XX
 PE 22-JUN-1998; 98MO-JP02765.
 XX
 PR 23-JUN-1997; 97JP-0165437.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
 XX
 DR WPI; 1999-105614/09.
 XX

PT Use of G protein-coupled receptor ligands - for modulating prolactin
 secretion or placental function, e.g. for treating menopausal
 syndrome, tumours, autoimmune disease or abnormal pregnancy
 XX
 PS Claim 3; Page 153; 241pp; English.
 CC The present sequence represents a rat type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypovarianism, gonocyst cacogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, emmenopathy, autoimmune disease,
 CC prolactinoma, infertility, impotence, amenorrhoea, galactorrhea,
 CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorioncarcinoma, hydatid mole, abortion, unthrifty fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
 CC
 SQ Sequence 31 AA;

Query Match 94.2%, Score 147; DB 20; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
 DB 1 SRAKHSMETRPDINPAMWTGKIRPVGRF 31

RESULT 4
 AAM97235
 ID AAM97235 standard; peptide; 31 AA.

XX AAM97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
 KW G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacogenesis;
 KW menopausal syndrome; euthyroid; hypometabolism; lactation;
 KW pituitary adenomatosis; brain tumour; emmenopathy; autoimmune disease;
 KW prolactinoma; infertility; impotence; amenorrhoea; galactorrhea;
 KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
 KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
 KW contraceptive; placental function; chorioncarcinoma; hydatid mole;
 KW abortion; unthrifty fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.

OS Homo sapiens.

XX W09858962-A1.

PN 30-DEC-1998

PD 22-JUN-1998; 96WO-JP02765.

PF 23-JUN-1997; 97JP-0165437.

PR (TAKE) TAKEDA CHEM IND LTD.

PA Fujii R, Hinuma S, Kawamata Y, Matsumoto H;

PI WPI; 1999-105614/09,

DR Use of G protein-coupled receptor ligands - for modulating prolactin

PT secretion or placental function, e.g. for treating menopausal
 syndrome, tumours, autoimmune disease or abnormal pregnancy
 XX
 PS Claim 3; Page 159; 241pp; English.

CC The present sequence represents a human type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypovarianism, gonocyst cacogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, emmenopathy, autoimmune disease,
 CC prolactinoma, infertility, impotence, amenorrhoea, galactorrhea,
 CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorioncarcinoma, hydatid mole, abortion, unthrifty fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
 CC
 SQ Sequence 31 AA;

Query Match 94.2%, Score 147; DB 20; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
 DB 1 STHRHSMEIRTPDINPAMVYASRGIRPVGRF 31

RESULT 5
 AAM87614
 ID AAM87614 standard; peptide; 31 AA.

XX AAM87614;

DT 29-MAR-1999 (first entry)

DE Rat 19P2 ligand.

XX 19P2 ligand; G protein coupled receptor; pituitary;
 KW prolactin releasing peptide; rat; dementia; breast cancer;
 KW therapy.

OS Rattus sp.

XX EP887417-A2.

PN 30-DEC-1998.

PD 25-JUN-1998; 98EP-0111725.

PF 27-JUN-1997; 97JP-0172118.

PR (TAKE) TAKEDA CHEM IND LTD.

PA Moriya T, Nishimura O, Suenaga M, Tanaka Y;

PI WPI; 1999-047884/05.

DR Producing a 19P2 pituitary G protein receptor ligand - by cleavage

PT of a fusion protein, useful for preventing and treating dementia,

PI breast cancer, renal failure and autoimmune disease

PS Claim 5; Page 34; 56pp; English.

CC This is the amino acid sequence of the rat pituitary G

CC protein-coupled receptor ligand 19P2L. A method suitable for

CC commercial high-level production of 19P2L comprises expressing

PT Physiologically-active polypeptide recognized as ligand by G
 CC protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 CC medicine
 XX
 PS Claim 3; Page 57; 72pp; Japanese.
 CC
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.,
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a rat peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 XX
 SQ Sequence 31 AA;
 CC
 Query Match 94.2%; Score 147; DB 21; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SRXHSHEKRTPDINPAMYXRGIRPVGRF 31
 DB 1 SRXHSHEKRTPDINPAMYXRGIRPVGRF 31
 XX
 RESULT 10
 AAB10362
 ID AAB10362 standard; peptide; 31 AA.
 XX
 AC AAB10362;
 XX
 DT 24-NOV-2000 (first entry)
 XX
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 32.
 XX
 KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.
 XX
 OS Homo sapiens.
 XX
 PN WO200038704-A1.
 XX
 PD 06-JUL-2000.
 XX
 PF 22-DEC-1999; 99WO-JP07199.
 XX
 PR 25-DEC-1998; 98JP-0369585.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Matsumoto H, Kitada C, Hinuma S;
 XX
 DR WPI, 2000-452298/39.
 XX
 PT Physiologically-active polypeptide recognized as ligand by G
 CC protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 CC medicine
 XX
 PS Disclosure; Page 62; 72pp; Japanese.
 CC
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.,
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial

CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 XX
 SQ Sequence 31 AA;
 CC
 Query Match 94.2%; Score 147; DB 21; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SRXHSHEKRTPDINPAMYXRGIRPVGRF 31
 DB 1 SRXHSHEKRTPDINPAMYXRGIRPVGRF 31
 XX
 RESULT 11
 AAY87504
 ID AAY87504 standard; protein; 31 AA.
 XX
 AC AAY87504;
 XX
 DT 18-JUL-2000 (first entry)
 XX
 DE Rat prolactin-releasing peptide, PrRP.
 XX
 KW Prolactin-releasing peptide; PrRP; GPR10; G protein-coupled receptor;
 KW feeding behaviour; food intake; modulation; antagonist; anorectic;
 KW obesity; agonist; cachexia.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 31
 FT /note="C-terminal amide"
 XX
 PN WO200017641-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 22-SEP-1999; 99WO-US21243.
 XX
 PR 22-SEP-1998; 98US-0101380.
 PR 14-OCT-1998; 98US-0172353.
 XX
 PA (MILL-) MILLENNium PHARM INC.
 XX
 PI Stricker-Kongrad A, Gu W;
 XX
 DR WPI, 2000-303231/26.
 XX
 PT Identifying modulators of body weight by a combination of a cell-free
 CC or cell-based assay to identify modulators of GPR10, followed by an in
 CC vivo assay for the compounds effect on e.g. feeding behavior
 XX
 PS Example 2; Page 61; 82pp; English.
 CC
 CC The invention relates to a method for identifying compounds useful for
 CC modulating body weight. The method comprises cell-free and/or cell-based
 CC assays that identify compounds which bind to and/or activate or inhibit
 CC the activity of GPR10, a G protein-coupled receptor. These assays are
 CC then followed by an in vivo assay of the effect of the compound on
 CC feeding behaviour, body weight or metabolic rate in a mammal. Prolactin-
 CC releasing peptide (PrRP; AAY87504) is a ligand of GPR10. Binding of PrRP
 CC to GPR10 stimulates a signal transduction cascade, which results in an
 CC increase in food intake. Compounds identified using the method of the
 CC invention are useful for the modulation of body weight. Antagonists of
 CC GPR10 can be used to treat obesity, while GPR10 agonists can be used to
 CC treat cachexia. The present sequence represents rat PrRP.
 XX
 SQ Sequence 31 AA;
 CC
 Query Match 94.2%; Score 147; DB 21; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.7e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

RESULT 12

AAV49291
ID AAV49291 standard; peptide; 31 AA.

AC AAV49291;

DE 22-FEB-2000 (first entry)

DE 19P2 ligand peptide fragment.

XX Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.
XX Homo sapiens.

OS

Key Location/Qualifiers
Modified-site 31 /note="C-terminal amide"

PN W09960112-A1.

PD 25-NOV-1999.

PF 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-039381/03.

PT New monoclonal antibodies, useful in diagnosis, as drugs and in
studying diseases related to ligand abnormality

PS Disclosure; Page 26; 73pp; Japanese.

XX The invention provides a monoclonal antibody which has a specific
reaction with the part peptide of the C-terminal of 19P2 ligand or its
derivative. The antibodies can be used in diagnosis or to treat or
prevent diseases associated with abnormality in the pituitary function
regulatory mechanism (e.g. promotion of prolactin secretion), central
nervous regulatory mechanism, and pancreatic function regulatory
mechanism. The antibody-based immunosay can also be applied in
clarifying the physiological functions of the ligand and its derivative.
XX Sequences AAV49290-302 represent peptide fragments of the 19P2 ligand.

SO Sequence 31 AA;

Query Match 94.2%; Score 147; DB 21; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.7e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

RESULT 13

AAV49292
ID AAV49292 standard; peptide; 31 AA.

AC AAV49292;

XX

DT 22-FEB-2000 (first entry)

DE 19P2 ligand peptide fragment.

XX Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.
XX Rattus sp.

OS

Key Location/Qualifiers
Modified-site 31 /note="C-terminal amide"

PN W09960112-A1.

PD 25-NOV-1999.

PF 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-039381/03.

PT New monoclonal antibodies, useful in diagnosis, as drugs and in
studying diseases related to ligand abnormality

PS Disclosure; Page 26; 73pp; Japanese.

XX The invention provides a monoclonal antibody which has a specific
reaction with the part peptide of the C-terminal of 19P2 ligand or its
derivative. The antibodies can be used in diagnosis or to treat or
prevent diseases associated with abnormality in the pituitary function
regulatory mechanism (e.g. promotion of prolactin secretion), central
nervous regulatory mechanism, and pancreatic function regulatory
mechanism. The antibody-based immunosay can also be applied in
clarifying the physiological functions of the ligand and its derivative.
XX Sequences AAV49290-302 represent peptide fragments of the 19P2 ligand.

SO Sequence 31 AA;

Query Match 94.2%; Score 147; DB 21; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.7e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

RESULT 14

AAV62524
ID AAV62524 standard; peptide; 31 AA.

AC AAV62524;

DE 24-AUG-2001 (first entry)

DE Rat CRH releasing protein related peptide SEQ ID NO: 18.

XX Rat; corticotrophin releasing hormone; CRH; G protein receptor ligand;
KW analgesic; hypercortisolemia; hypercortisolemia; hypoadrenocorticism;
Addison's disease; adrenal gland hyperfunction; obesity.

OS Rattus sp.

PN W0200135984-A1.

PD 25-MAY-2001.

XX

```

PF 17-NOV-2000; 2000WO-JP08119.
XX
XX 18-NOV-1999; 99JP-0327900.
PR 26-SEP-2000; 2000JP-0297073.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
PA
XX Kitada C, Matsumoto H, Hinuma S;
XX WPI; 2001-355552/37.
DR
XX Use of G protein receptor ligand or peptide for controlling
PT corticotrophin releasing hormone secretion -
XX
XX Claim 3; page 69; 90pp; Japanese.
PS
XX The present sequence describes a method of controlling the secretion of
CC corticotrophin releasing hormone (CRH), involving the use of a G protein
CC receptor ligand. This can be used to control the secretion of CRH and is
CC useful as an analgesic or for treating, preventing or ameliorating
CC diseases associated with CRH secretion such as hyperaldosteronism,
CC Addison's disease (including boredom, nausea, pigmentation, hypogonadism, hair
CC loss, and hypotension), adrenal gland hypofunction and obesity. The
CC present sequence is a peptide used in the exemplification of the
CC invention.
XX
SQ Sequence 31 AA;
Query Match 94.2%; Score 147; DB 22; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.7e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SRHXHSMEXRTPDINPAWYXXRGIRPYGRF 31
DB 1 SRAQHSMEIRTPDINPAWYTGKINPYGRF 31

RESULT 15
AAG62531
ID AAG62531 standard; peptide; 31 AA.
XX
XX AAG62531;
AC
XX 24-AUG-2001 (first entry)
XX
XX Human CRH releasing protein related peptide SEQ ID NO: 32.
DE
XX Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
XX analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
XX Addison's disease; adrenal gland hypofunction; obesity.
XX
XX Homo sapiens.
OS
XX WO200135984-A1.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-JP08119.
XX
XX 18-NOV-1999; 99JP-0327900.
XX
XX 26-SEP-2000; 2000JP-0297073.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX Kitada C, Matsumoto H, Hinuma S;
XX WPI; 2001-355552/37.
XX
XX Use of G protein receptor ligand or peptide for controlling
XX corticotrophin releasing hormone secretion -
XX
XX Claim 3; page 73-74; 90pp; Japanese.
PS

```

```

XX
XX The present sequence describes a method of controlling the secretion of
CC corticotrophin releasing hormone (CRH), involving the use of a G protein
CC receptor ligand. This can be used to control the secretion of CRH and is
CC useful as an analgesic or for treating, preventing or ameliorating
CC diseases associated with CRH secretion such as hyperaldosteronism,
CC Addison's disease (including boredom, nausea, pigmentation, hypogonadism, hair
CC loss, and hypotension), adrenal gland hypofunction and obesity. The
CC present sequence is a peptide used in the exemplification of the
CC invention.
XX
SQ Sequence 31 AA;
Query Match 94.2%; Score 147; DB 22; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.7e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SRHXHSMEXRTPDINPAWYXXRGIRPYGRF 31
DB 1 SRAQHSMEIRTPDINPAWYASRGIRPYGRF 31

Search completed: May 1, 2003, 14:46:48
Job time : 35 secs

```

Thu May 1 15:54:14 2003

de-lun

us-09-446-543a-61.rag

SEE ID NO: 61
Ae NO: AAW31391

datafile: A-Geneseq-101002

File Copy
Page 1
Cited in paper #

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 : Search time 43.1094 Seconds
(without alignments)
95.821 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171

Sequence: 1 SRHRHSMETRPDIPNPNVYASRGIPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

- 1: /SIDS2/gcgdata/geneseq/emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
V1	171	100.0	31	18	AAW31391
2	171	100.0	31	20	AAW7233
3	171	100.0	31	20	AAW87615
4	171	100.0	31	21	AAW10362
5	171	100.0	31	21	AAV49291
6	171	100.0	31	22	AAW62531
7	171	100.0	31	22	AAW90991
8	171	100.0	32	18	AAW31392
9	171	100.0	32	21	AAW10363
10	171	100.0	32	22	AAW62532

11	171	100.0	33	18	AAW31393	Human type G prote
12	171	100.0	33	21	AAW10364	Human oxytocin sec
13	171	100.0	33	22	AAW62533	Human CRH releasin
14	171	100.0	87	18	AAW31390	Human type G prote
15	171	100.0	87	20	AAW97226	Human type G prote
16	171	100.0	87	21	AAW10361	Human oxytocin sec
17	171	100.0	87	21	AAW62530	Human CRH releasin
18	165	96.5	30	21	AAV49299	19P2 ligand peptid
19	162	94.7	31	22	AAW90995	Proctatin releasin
20	158	92.4	31	18	AAW31371	Bovine G protein-c
21	158	92.4	31	20	AAW97218	Bovine pituitary-d
22	158	92.4	31	20	AAW87613	Bovine 19P2 ligand
23	158	92.4	31	20	AAW95188	Bovine pituitary-d
24	158	92.4	31	21	AAW10347	Bovine oxytocin se
25	158	92.4	31	21	AAW49290	19P2 ligand peptid
26	158	92.4	31	21	AAV49298	19P2 ligand peptid
27	158	92.4	31	22	AAW62516	Bovine CRH releas
28	158	92.4	32	18	AAW31372	Bovine G protein-c
29	158	92.4	32	20	AAW95189	Bovine pituitary-d
30	158	92.4	32	21	AAW10348	Bovine oxytocin se
31	158	92.4	32	22	AAW62517	Bovine CRH releas
32	158	92.4	33	16	AAW31373	Bovine genome deri
33	158	92.4	33	20	AAW95190	Bovine G protein-c
34	158	92.4	33	21	AAW10349	Bovine pituitary-d
35	158	92.4	33	21	AAV49297	Bovine oxytocin se
36	158	92.4	33	22	AAW62518	19P2 ligand peptid
37	158	92.4	98	18	AAW31382	Bovine CRH releas
38	158	92.4	98	18	AAW31366	Bovine genome deri
39	158	92.4	98	20	AAW97224	Bovine G protein-c
40	158	92.4	98	20	AAW97217	Bovine genome-driv
41	158	92.4	98	20	AAW95187	Bovine pituitary-d
42	158	92.4	98	21	AAW10346	Bovine genome-deri
43	158	92.4	98	21	AAW10353	Bovine oxytocin se
44	158	92.4	98	22	AAW62515	Bovine CRH releas
45	158	92.4	98	22	AAW62522	Bovine CRH releas

ALIGNMENTS

RESULT 1
ID AAW31391 standard, Peptide: 31 AA.
AC AAW31391:
XX 06-APR-1998 (first entry)
DT
XX Human type G protein-coupled receptor ligand fragment 1.
XX DE
XX G protein-coupled receptor; ligand binding; pharmaceutical;
XX KW modulator; pituitary; central nervous system; pancreas; prophylactic;
XX KW therapeutic agent.
XX
XX Homo sapiens.
XX PN WO9724436-A2
XX 10-JUL-1997.
XX PD
XX 26-DEC-1996: 96W0-JP03821.
XX PF
XX 18-SEP-1996: 96JP-0246573.
XX PR 28-DEC-1995: 95JP-0343371.
XX PR 15-MAR-1996: 96JP-0059419.
XX PR 12-AUG-1996: 96JP-0211805.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX Fujii R, Fukushima S, Habata Y, Hinuma S, Hosoya M,
XX Kawamata Y, Kitada C,
XX WPI, 1997-363672/33.

DR N-PSDB: AAV02428.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2; Page 184; 258pp; English.

CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 23 to 53 of the
 CC sequence represented in AAW1390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator. A central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligogalacta. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.

SQ Sequence 31 AA;

Query Match 100.0%; Score 171; DB 18; Length 31;
 Best Local Similarity 100.0%; Pred. No. 4.9e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMELRTPDINPAWYASRGIRPVGRF 31
 |||
 DB 1 SRTHRSMELRTPDINPAWYASRGIRPVGRF 31

RESULT 2
 AAW97235
 ID AAW97235 standard; peptide; 31 AA.

XX AAW97235;
 AC
 XX
 DT 06-MAY-1999 (first entry)
 XX

DE Human type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
 KW G protein-coupled receptor; GPCR; hypocoarctinism; gonocyst cacogenesis;
 KW menopausal syndrome; euthyroid; hypometabolism; lactation;
 KW pituitary adenomatosis; brain tumour; amenorrhoea; galactorrhea;
 KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
 KW acromegaly; Chlari-Frommel syndrome; Argonz-del Castillo syndrome;
 KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
 KW contraceptive; placental function; chorionicarctinoma; hydatid mole;
 KW interruption mole; abortion; unfertility fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.

XX Homo sapiens.

XX OS

XX W09858962-A1.

XX 30-DEC-1998.

XX 22-JUN-1998; 98WO-JP02765.

XX 23-JUN-1997; 97JP-0165437.

XX (TAKE) TAKEDA CHEM IND LTD.
 XX Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
 XX

DR WPI: 1999-105614/09.

XX Use of G protein-coupled receptor ligands - for modulating prolactin
 PT secretion or placental function, e.g. for treating menopausal
 PT syndrome, tumours, autoimmune disease or abnormal pregnancy
 XX
 PS Claim 3; Page 159; 244pp; English.

CC The present sequence represents a human type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypocoarctinism, gonocyst cacogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, amenorrhoea, galactorrhea,
 CC prolactinoma, infertility, impotence, Argonz-del Castillo syndrome,
 CC acromegaly, Chlari-Frommel syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorionicarctinoma, hydatid mole, interruption mole, unfertility fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

SQ Sequence 31 AA;

Query Match 100.0%; Score 171; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 4.9e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMELRTPDINPAWYASRGIRPVGRF 31
 |||
 DB 1 SRTHRSMELRTPDINPAWYASRGIRPVGRF 31

RESULT 3
 AAW87615
 ID AAW87615 standard; peptide; 31 AA.

XX AAW87615;
 AC
 XX
 DT 29-MAR-1999 (first entry)
 XX

DE Human 1992 ligand.

XX 1992 ligand; G protein coupled receptor; pituitary;
 KW prolactin releasing peptide; human; dementia; breast cancer;
 KW therapy.

XX Homo sapiens.

XX EP867417-A2.

XX 30-DEC-1998.

XX 25-JUN-1998; 98EP-0111725.

XX 27-JUN-1997; 97JP-0172118.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Moriya T, Nishimura O, Suenaga M, Tanaka Y;

XX WPI: 1999-047884/05.

XX Producing a 1992 pituitary G protein receptor ligand - by cleavage
 PT of a fusion protein, useful for preventing and treating dementia,
 PT breast cancer, renal failure and autoimmune disease
 XX
 PS Claim 5; Page 35; 56pp; English.

This is the amino acid sequence of the human pituitary G protein-coupled receptor ligand 19p2L. A method suitable for commercial high-level production of 19p2L comprises expressing the ligand in host cells as a recombinant fusion protein e.g. with human basic fibroblast growth factor (see AAV83796-97) that has been modified to include an N-terminal cysteine residue. The ligand is released from the fusion by cyanidation followed by ammonolysis. 19p2L has prolactin secretion-stimulating and (at high doses) prolactin secretion-inhibiting properties. It can be used in the treatment and prevention of various diseases including: senile dementia, cerebrovascular dementia, and dementia associated with: neurological disorders (e.g. Alzheimer's disease, Parkinson's disease, Pick's disease, Huntington's disease), infectious diseases (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism, intoxication by drugs, metal and organic compounds), tumorigenic diseases (e.g. brain tumour), traumatic diseases (e.g. chronic subarachnoid haemorrhage, and other types of dementia, depression, hyperactive child syndrome (microencephalopathy) and disturbance of consciousness. It is also useful for prevention and treatment of diseases associated with prolactin hypo and hypersecretion respectively, including: hyperprolactinaemia, pituitary adenoma, breast cancer, infertility, impotence and autoimmune disease (hypersecretion disorders), and seminal vesicle hypoplasia, osteoporosis, menopausal syndrome and renal failure (hyposecretion disorders). The 19p2 polypeptide/amide is also useful as a test reagent for study of the prolactin secretory function or as a lactagogue in mammalian farm animals.

Sequence 31 AA;

Query Match 100.0%; Score 171; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 4, 9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTSRHSMEIRTPDINPAMWASRGIRPVGRF 31
|||||
DB 1 SRTSRHSMEIRTPDINPAMWASRGIRPVGRF 31

RESULT 4
AAB10362
ID AAB10362 standard; peptide: 31 AA.

AC AAB10362;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 32.

KM Human: oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KM caesarean section; artificial fertilization; galactostasis; goat; pig;
KM veterinary medicine; milk production.

OS Homo sapiens.

PN WO200038704-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

PT Physiologically-active polypeptide recognized as ligand by G
protein-coupled receptor protein, for promoting secretion of oxytocin.

PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine
XX
PS Disclosure; Page 62; 72pp; Japanese.

CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a human peptide which acts as an oxytocin secretion
CC promoter.

Sequence 31 AA;

Query Match 100.0%; Score 171; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4, 9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTSRHSMEIRTPDINPAMWASRGIRPVGRF 31
|||||
DB 1 SRTSRHSMEIRTPDINPAMWASRGIRPVGRF 31

RESULT 5
AAV49291
ID AAV49291 standard; peptide: 31 AA.

AC AAV49291;

DT 22-FEB-2000 (first entry)

DE 19p2 ligand peptide fragment.

KM Monoclonal antibody; 19p2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.
OS Homo sapiens.

FT Key Location/Qualifiers
FT Modified-site 31
/note- "C-terminal amide"

PN WO960112-A1.

PD 25-NOV-1999.

PF 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-039381/03.

PT New monoclonal antibodies, useful in diagnosis, as drugs and in
PT studying diseases related to ligand abnormality
XX
PS Disclosure; Page 26; 73pp; Japanese.

CC The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19p2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion), central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in
CC clarifying the physiological functions of the ligand and its derivative.

XX Human type G protein-coupled receptor ligand fragment 2.
 KW G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.
 XX
 OS Homo sapiens.
 XX
 PN WO9724436-A2.
 XX
 PD 10-JUL-1997.
 XX
 PE 26-DEC-1996; 96WO-JP03821.
 XX
 PR 18-SEP-1996; 96JP-0246573.
 PR 28-DEC-1995; 95JP-0343371.
 PR 15-MAR-1996; 96JP-0059419.
 PR 12-AUG-1996; 96JP-0211805.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Fukusumi S, Nabata Y, Hinuma S, Hosoya M;
 PI Kawamata Y, Kitada C;
 XX
 DR WPI: 1997-363672/33.
 DR N-PSDB; AAV02429.
 XX
 PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2; Page 185; 258pp; English.
 XX
 CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 23 to 54 of the
 CC sequence represented in AAW31390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC propylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, hyperkinetic
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia, diabetes,
 CC Turner's syndrome, neuritis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharide. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 CC
 XX
 SO Sequence 32 AA;
 Query Match 100.0%; Score 171; DB 18; Length 32;
 Best Local Similarity 100.0%; Pred. No. 5.1e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31
 DB 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31
 RESULT 9
 ID AAB10363 standard; peptide; 32 AA.
 AC AAB10363;
 XX
 XX 24-NOV-2000 (first entry)
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 33.

XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.
 XX
 OS Homo sapiens.
 XX
 PN WO200038704-A1.
 XX
 PD 06-JUL-2000.
 XX
 PE 22-DEC-1999; 99WO-JP07199.
 XX
 PR 25-DEC-1998; 98JP-0369585.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Matsumoto H, Kitada C, Hinuma S;
 XX
 DR WPI: 2000-452298/39.
 XX
 PT Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine
 XX
 PS Disclosure; Page 62; 72pp; Japanese.
 XX
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 XX
 SO Sequence 32 AA;
 Query Match 100.0%; Score 171; DB 21; Length 32;
 Best Local Similarity 100.0%; Pred. No. 5.1e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31
 DB 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31
 RESULT 10
 ID AAG62532 standard; peptide; 32 AA.
 AC AAG62532;
 XX
 XX 24-AUG-2001 (first entry)
 DE Human CRH releasing protein related peptide SEQ ID NO: 33.
 KW Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
 KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
 KW Addison's disease; adrenal gland hyperfunction; obesity.
 XX
 OS Homo sapiens.
 XX
 PN WO200135984-A1.
 XX
 PD 25-MAY-2001.
 XX
 DE 17-NOV-2000; 2000WO-JP08119.

PR 18-NOV-1999; 99JP-0327900.
 PR 26-SEP-2000; 2000JP-0297073.
 PA (TAKE) TAKEDA CHEM IND LTD.
 PI Kitada C, Matsumoto H, Hinuma S;
 DR WPI; 2001-355552/37.
 XX
 PR Use of G protein receptor ligand or peptide for controlling
 PT corticotrophin releasing hormone secretion .
 PS
 XX Disclosure; Page 74; 90pp; Japanese.
 CC The present sequence describes a method of controlling the secretion of
 CC corticotrophin releasing hormone (CRH), involving the use of a G protein
 CC receptor ligand. This can be used to control the secretion of CRH and is
 CC useful as an analgesic or for treating, preventing or ameliorating
 CC diseases associated with CRH secretion such as hyperaldosteronism,
 CC hypercortisolism, secondary or chronic hypoadrenocorticism, Addison's
 CC disease (including boredom, nausea, pigmentation, hypogonadism, hair
 CC loss, and hypotension), adrenal gland hypofunction and obesity. The
 CC present sequence is a peptide used in the exemplification of the
 CC invention.
 SQ Sequence 32 AA:
 Query Match 100.0%; Score 171; DB 22; Length 32;
 Best Local Similarity 100.0%; Pred. No. 5.1e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRTSHSMETRTPDINPAMYASRGIRPVGRF 31
 DB 1 SRTSHSMETRTPDINPAMYASRGIRPVGRF 31
 ID AAW31393 standard; Peptide; 33 AA.
 AC AAW31393;
 DT 06-APR-1998 (first entry)
 DE Human type G protein-coupled receptor ligand fragment 3.
 XX
 KW G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.
 OS Homo sapiens.
 XX
 PN WO9724436-A2.
 PD 10-JUL-1997.
 PF 26-DEC-1996; 96WO-JP03821.
 PR 18-SEP-1996; 96JP-0246573.
 PR 28-DEC-1995; 95JP-0343371.
 PR 15-MAR-1996; 96JP-0059419.
 PR 12-ADG-1996; 96JP-0211805.
 PA (TAKE) TAKEDA CHEM IND LTD.
 PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
 PI Kawamata Y, Kitada C;
 DR WPI; 1997-363672/33.
 DR N-PSDB; AAW02430.
 XX
 PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland

XX Claim 2; Page 185; 258pp; English.
 PS
 XX This sequence represents a peptide fragment from a novel human type
 CC sequence represented in AAW31390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 SQ Sequence 33 AA:
 Query Match 100.0%; Score 171; DB 18; Length 33;
 Best Local Similarity 100.0%; Pred. No. 5.3e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRTSHSMETRTPDINPAMYASRGIRPVGRF 31
 DB 1 SRTSHSMETRTPDINPAMYASRGIRPVGRF 31
 ID AAB10364 standard; peptide; 33 AA.
 AC AAB10364;
 DT 24-NOV-2000 (first entry)
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 34.
 XX
 KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.
 OS Homo sapiens.
 XX
 PN WO200036704-A1.
 PD 06-JUL-2000.
 PF 22-DEC-1999; 99WO-JP07199.
 PR 25-DEC-1998; 98JP-0369585.
 PA (TAKE) TAKEDA CHEM IND LTD.
 PI Matsumoto H, Kitada C, Hinuma S;
 DR WPI; 2000-452298/39.
 PT Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine .
 PS Disclosure; Page 62-63; 72pp; Japanese.
 XX This invention describes a novel oxytocin secretion-regulating agent

CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.

XX
 XX Sequence 33 AA;

Query Match Best Local Similarity 100.0%; Score 171; DB 21; Length 33;
 Matched 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTSRHSMEIRTPDINPAMVYASRGIRPVGRF 31
 DB 1 SRTSRHSMEIRTPDINPAMVYASRGIRPVGRF 31

RESULT 13

ID AAG62533 standard; peptide; 33 AA.

XX AAG62533;

XX 24-AUG-2001 (first entry)

DE Human CRH releasing protein related peptide SEQ ID NO: 34.

XX Human; corticotrophin releasing hormone; CRH: G protein receptor ligand;
 KW analgesic; hyperaldosteronism; hypercortisolism; hypodrenocorticism;
 KH Addison's disease; adrenal gland hypertrophy; obesity.

OS Homo sapiens.

XX W0200135984-A1.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000MO-JP08119.

XX 18-NOV-1999; 99JP-0327900.

XX 26-SEP-2000; 2000JP-0297073.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Kikada C, Matsumoto H, Hinuma S;

XX WPI; 2001-355552/37.

XX Use of G protein receptor ligand or peptide for controlling

XX corticotrophin releasing hormone secretion.

XX Disclosure; Page 74; 90pp; Japanese.

XX The present sequence describes a method of controlling the secretion of
 CC corticotrophin releasing hormone (CRH), involving the use of a G protein
 CC receptor ligand. This can be used to control the secretion of CRH and is
 CC useful as an analgesic or for treating, preventing or ameliorating
 CC diseases associated with CRH secretion such as hyperaldosteronism,
 CC hypercortisolism, secondary or chronic hypoadrenocorticism, Addison's
 CC disease (including bledom, nausea, pigmentation, hypogonadism, hair
 CC loss, and hypotension), adrenal gland hypofunction and obesity. The
 CC present sequence is a peptide used in the exemplification of the
 CC invention.

XX Sequence 33 AA;

Query Match Best Local Similarity 100.0%; Score 171; DB 22; Length 33;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTSRHSMEIRTPDINPAMVYASRGIRPVGRF 31
 DB 1 SRTSRHSMEIRTPDINPAMVYASRGIRPVGRF 31

RESULT 14

ID AAW31390 standard; protein; 87 AA.

XX AAW31390;

XX 06-APR-1998 (first entry)

DE Human type G protein-coupled receptor ligand encoded by PHO7.

XX G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.

XX Homo sapiens.

XX W09724436-A2.

XX 10-JUL-1997.

XX 26-DEC-1996; 96MO-JP03821.

XX 18-SEP-1996; 96JP-0246573.

XX 28-DEC-1995; 95JP-0343371.

XX 15-MAR-1996; 96JP-0059419.

XX 12-AUG-1996; 96JP-0211805.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;

XX Kawamata Y, Kikada C;

XX WPI; 1997-363672/33.

XX N-PSDB; AAV02427.

XX Ligand peptide for G protein-coupled receptor - acts by modulating

XX function in the central nervous system, pancreas and pituitary gland

XX Claim 3; Page 183; 258pp; English.

XX This sequence represents a novel human type ligand polypeptide encoded
 CC by PHO7 which is used in an assay to monitor ligand binding to the G
 CC protein-coupled receptor protein. Pharmaceutical compositions
 CC containing this ligand may be used as a pituitary function modulator, a
 CC central nervous system modulator or a pancreatic function modulator. This
 CC ligand could have specific applications as a prophylactic or therapeutic
 CC agent for dementia, depression, hyperkinetic syndrome, disturbance of
 CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
 CC secretory disease, hyperlipidemia, hypercholesterolemia,
 CC hyperglycemia, hyperlipidemia, hyperproliferative diseases,
 CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
 CC rheumatoid arthritis, spinal injury, transverse brain ischemia,
 CC amyotrophic lateral sclerosis, acute myocardial infarction,
 CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
 CC osteoporosis, asthma, epilepsy, infertility and/or oligosaccharide.
 CC can also be developed to screen compounds which are capable of altering
 CC the binding activity of the ligand thus affecting activation of the G
 CC protein-coupled receptor protein.

XX Sequence 87 AA;

Query Match Best Local Similarity 100.0%; Score 171; DB 18; Length 87;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 23 SRTHRHSEIRTPDINPAWYASRGIRPVGRF 53

RESULT 15
AAM97226

ID AAM97226 standard; peptide; 87 AA.

AC AAM97226;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide.

KW G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacosgenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation; modulation;
KW pituitary adenomatosis; brain tumour; amenorrhoea; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhoea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydralid mole;
KW interruption mole; abortion; unfertilized fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia; prolactin secretion;
KW human type ligand.

OS Homo sapiens.

PN W09858962-A1.

PD 30-DEC-1998.

PE 22-JUN-1998; 98WO-JP02765.

PR 23-JUN-1997; 97JP-0165437.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;

DR WPI: 1999-105614/09.

N-PSDB; AAX15526.

PT Use of G protein-coupled receptor ligands - for modulating prolactin
secretion or placental function, e.g. for treating menopausal
syndrome, tumours, autoimmune disease or abnormal pregnancy

PS Disclosure; Page 158; 241pp; English.

XX The present sequence represents a human type ligand polypeptide. The
XX specification describes an agent for modulating prolactin secretion
XX which comprises a ligand polypeptide or a salt, for a G protein-coupled
XX receptor (GPCR) protein. The agents for promoting prolactin secretion
XX can be used for treating or preventing hypovarianism, gonocyst
XX cacosgenesis, menopausal syndrome, euthyroid or hypometabolism. They
XX can be used for promoting lactation in a domestic mammal and as an
XX aphrodisiac. The agents for inhibiting prolactin secretion can be used
XX for treating or preventing pituitary adenomatosis, brain tumour,
XX amenorrhoea, autoimmune disease, prolactinoma, infertility, impotence,
XX acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
XX Forbes-Albright syndrome, lymphoma, Sheehan syndrome, dyszoospermia,
XX or dyszoospermia. The inhibitory agents can also be used as
XX contraceptives. The agents for modulating placental function can be used
XX for treating or preventing choriocarcinoma, hydralid mole, interruption
XX mole, abortion, unfertilized fetus, abnormal saccharometabolism, abnormal
XX lipidmetabolism or oxytocia.

SO Sequence 87 AA;

Query Match 100.0%; Score 171; DB 20; Length 87;

Best Local Similarity 100.0%; Pred. No. 1,7e-18;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 SRTHRHSEIRTPDINPAWYASRGIRPVGRF 31
|||||

DB 23 SRTHRHSEIRTPDINPAWYASRGIRPVGRF 53

Search completed: May 1, 2003, 14:30:49
Job time : 44.1094 secs

Thu May 1 15:54:25 2003

us-09-446-543a-73.rsp

Page 1

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 7.21875 Seconds

(without alignments)
126,404 Million cell updates/sec

Title: US-09-446-543A-73

Perfect score: 109

Sequence: 1 TPDIINPMAYXARGLRPGVGRFX 22

Scoring table: BIOSRM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	87	1	PRRP_HUMAN
2	104	95.4	83	1	PRRP_RAT
3	104	95.4	98	1	PRRP_BOVIN
4	46	42.2	676	1	EXIL_HUMAN
5	43	39.4	376	1	OPSL_LIMPO
6	43	39.4	376	1	OPSL_LIMPO
7	43	39.4	376	1	OPSL_LIMPO
8	43	39.4	376	1	OPSL_LIMPO
9	43	39.4	376	1	OPSL_LIMPO
10	43	39.4	376	1	OPSL_LIMPO
11	41.5	38.1	402	1	EXTL_SYRGO
12	41	37.6	342	1	ARGC_BACHD
13	41	37.6	342	1	Y762_METUA
14	41	37.6	342	1	Y762_METUA
15	40	36.7	329	1	Y493_MYCTU
16	40	36.7	329	1	Y493_MYCTU
17	40	36.7	329	1	Y493_MYCTU
18	40	36.7	329	1	Y493_MYCTU
19	40	36.7	329	1	Y493_MYCTU
20	40	36.7	329	1	Y493_MYCTU
21	40	36.7	329	1	Y493_MYCTU
22	40	36.7	329	1	Y493_MYCTU
23	39.5	36.2	860	1	VG12_BPB03
24	39	35.8	149	1	ENRN_BPT7
25	39	35.8	360	1	ENRN_BPT7
26	39	35.8	377	1	OPSL_HUMAN
27	39	35.8	377	1	OPSL_HUMAN
28	39	35.8	377	1	OPSL_HUMAN
29	39	35.8	377	1	OPSL_HUMAN
30	39	35.8	377	1	OPSL_HUMAN
31	39	35.8	377	1	OPSL_HUMAN
32	38.5	35.3	240	1	OBIE_RICCN
33	38.5	35.3	248	1	OBIE_RICCN

34	38.5	35.3	248	1	OBIE_RICCN	09263 rickettsia
35	38.5	35.3	831	1	DPF_ETHH	P30313 thermus the
36	38.5	35.3	877	1	CAD2_BOVIN	P19534 bos taurus
37	38.5	35.3	906	1	CAD2_HUMAN	P19022 homo sapien
38	38.5	35.3	906	1	CAD2_MOUSE	P15116 mus musculu
39	38.5	35.3	906	1	CAD2_MOUSE	092193 ratu mus
40	38.5	35.3	1083	1	POLG_ZYMR	089310 z genome po
41	38	34.9	158	1	R1L5_AERPE	09198 aeropyrum p
42	38	34.9	197	1	118B_HUMAN	095998 homo sapien
43	38	34.9	261	1	PBSC_ECOLI	P77404 escherichia
44	38	34.9	288	1	Y587_PASMU	09265 pasteurella
45	38	34.9	332	1	ISPH_MYCLE	092781 mycobacteri

ALIGNMENTS

RESULT 1
ID PRRP_HUMAN STANDARD, PRT; 87 AA.
AC P81277
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PRPP)
DE hormone) (Contains: Proactin-releasing peptide PRPP31; Proactin-
DE releasing peptide PRPP20).
GN PRR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96266781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.,
RA "A proactin-releasing peptide in the brain."
RL Nature 393:272-276(1998).
RT [2]
RN TISSUE SPECIFICITY.
RX MEDLINE=99426652; PubMed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S.,
RA Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H.,
RA Sumino Y., Fujino M.,
RT "Tissue distribution of proactin-releasing peptide (PRPP) and its
RT receptor."
RL Regul. Pept. 83:1-10(1999).
CC -!- FUNCTION: Stimulates prolactin (PRL) release and regulates the
CC expression of prolactin through its receptor GPR10. May stimulate
CC lactotrophs directly to secrete PRL.
CC -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AB015419; BAA29027.1;
CC DR MTM; 602663;
CC KW Hormone; Amidation; Signal.
CC FT SIGNAL. 1
CC FT DEPTIDE 23 53
CC FT PEPTIDE 34 53
CC FT MOD_RES 53 53
CC SO SEQUENCE 87 AA; 9639 MW; 229A2F350CF981B CRC64;
Query Match 96.3%; Score 105; DB 1; Length 87;

Best Local Similarity 90.0%; Pred. No. 1, Le-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXNGIRPVGRF 20
Db 34 TPDINPAMYXXNGIRPVGRF 53

RESULT 2

PRRP_RAT STANDARD; PRT; 83 AA.

AC P81278;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Prolactin-releasing peptide precursor (PRP) (Prolactin-releasing hormone) [Contains: Prolactin-releasing peptide PRRP31; Prolactin-releasing peptide PRRP20].
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE-BRAIN;
RC MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain."
RL Nature 393:272-276(1998).
RN [2]
RN TISSUE SPECIFICITY:
RX MEDLINE=99426652; PubMed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H., Sumino Y., Fujino M.;
RT "Tissue distribution of prolactin-releasing peptide (PRP) and its receptor."
RL Regul. Pept. 83:1-10(1999).
CC -1- FUNCTION: Stimulates prolactin (PR) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PR.
CC -1- TISSUE SPECIFICITY: Widely expressed, with highest levels in medulla oblongata and hypothalamus.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR EMBL, AB015417; BAA29026.1;
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 21
FT PEPTIDE 22 52
FT PEPTIDE 33 53
FT MOD RES 52 52
SO SEQUENCE 83 AA; 9215 MW; DQC75A264EEF429 CRC64;

Query Match 95.4%; Score 104; DB 1; Length 83;
Best Local Similarity 90.0%; Pred. No. 1, Le-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXNGIRPVGRF 20
Db 33 TPDINPAMYXXNGIRPVGRF 52

RESULT 3

PRRP_BOVIN STANDARD; PRT; 98 AA.

AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Prolactin-releasing peptide precursor (PRP) (Prolactin-releasing hormone) [Contains: Prolactin-releasing peptide PRRP31; Prolactin-releasing peptide PRRP20].
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX [1]
RN SEQUENCE FROM N.A. AND SEQUENCE OF 23-52.
RP TISSUE-BRAIN;
RC MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain."
RL Nature 393:272-276(1998).
RN [2]
RN TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
RX MEDLINE=99426652; PubMed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H., Sumino Y., Fujino M.;
RT "Tissue distribution of prolactin-releasing peptide (PRP) and its receptor GPR10. May stimulate lactotrophs directly to secrete PR."

CC -1- FUNCTION: Stimulates prolactin (PR) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PR.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR EMBL, AB015417; BAA29025.1;
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22
FT PEPTIDE 23 53
FT PEPTIDE 33 53
FT MOD RES 53 53
SO SEQUENCE 98 AA; 10544 MW; 0BAC35A1380F908 CRC64;

Query Match 95.4%; Score 104; DB 1; Length 98;
Best Local Similarity 90.0%; Pred. No. 1, Le-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXNGIRPVGRF 20
Db 34 TPDINPAMYXXNGIRPVGRF 53

RESULT 4

EX1L_HUMAN STANDARD; PRT; 676 AA.

AC Q92935;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exostosin-like 1 (EC 2.4.1.-) (Exostosin-L) (Multiple exostosin-like protein).
GN EX1L OR EXTL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=97189339; PubMed=9037597;
RA Wise C.A., Cline G.A., Massa H., Trak B.J., Lovett M.;
RT "Identification and localization of the gene for EXTL, a third member

DE	Lateral eye opsin.
DS	Limulus polyphemus (Atlantic horseshoe crab).
OC	Eurypterus; Metazoa; Arthropoda; Chelicerata; Neurostomata; Xiphosura;
CC	Limulidae; Limulus.
OX	NCSI_TaxID=6850;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	TISSUE=Lateral eye;
RX	MEDLINE=93317641; PubMed=8327495;
RA	Smith W.C., Price D.A., Greenberg R.M., Batteille B.-A.;
RT	"Opsins from the lateral eyes and ocelli of the horseshoe crab,
RL	Limulus polyphemus.";
Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).	
-1-	FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
Mediate vision. They consist of an apoprotein, opsin, covalently	
linked to cis-retinal.	
-1-	SUBCELLULAR LOCATION: Integral membrane protein.
-1-	TISSUE SPECIFICITY: LATERAL EYE.
-1-	PFM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
BE PHOSPHORYLATED (BY SIMILARITY).	
-1-	MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AT 520 NM.
-1-	SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
OPSIN SUBFAMILY.	

CC	This SWISS-Prot entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
DR	EMBL; L03791; AAA28273.1; -
DR	EXBL; L03781; AAA02498.1; -
DR	PIR; B48197; B48197.
DR	HSSP; P02699; LEDV.
DR	InterPro: IPR000276; GPCR_Rhodopsn.
DR	InterPro: IPR001760; Opsin.
DR	Pfam; PF00001; 7tm_1; 1.
DR	PRINTS; PR00237; GPCRRHODPSN.
DR	PROSITE; PS00237; G.PROTEIN.RECP_FL_1; 1.
DR	PROSITE; PS50262; G.PROTEIN.RECP_FL_2; 1.
DR	PROSITE; PS00238; OPSIN; 1.
KW	Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KM	Phosphorylation; g-protein coupled receptor.
FT	DOMAIN 1 46
FT	TRANSSEM 47 71
FT	DOKAIN 72 83
FT	TRANSSEM 84 108
FT	DOKAIN 109 123
FT	TRANSSEM 124 143
FT	DOKAIN 144 162
FT	TRANSSEM 163 186
FT	DOKAIN 187 210
FT	TRANSSEM 211 238
FT	DOKAIN 239 274
FT	TRANSSEM 275 298
FT	DOKAIN 299 306
FT	TRANSSEM 307 331
FT	DOKAIN 332 376
FT	DISULFID 120 197
FT	BINDING 318 318
FT	CARBOHYD 17 17
FT	CARBOHYD 193 193
SO	SEQUENCE 376 AA; 42139 MW; CCE401766A06F26 CRC64;
QY	Query Match 39.4%; Score 43; DB 1; Length 376;
Best Local Similarity 46.2%; Pred. No. 8.7;	
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;	
2 PDINPATYXRGI 14	
I::I I I	
40 PPNNPLVSLTGV 52	

FT BINDING 307 307 HEME 4 (COVALENT) (BY SIMILARITY).
 FT BINDING 310 310 HEME 4 (COVALENT) (BY SIMILARITY).
 FT METAL 311 311 IRON 4 (HEME AXIAL LIGAND)
 FT METAL 311 311 (BY SIMILARITY).
 SQ SEQUENCE 383 AA; 41522 MW; 96BCD91FF1B9AE7E CRC64;
 Query Match 39.4%; Score 43; DB 1; Length 383;
 Best Local Similarity 36.8%; Pred. No. 8.9;
 Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
 QY 1 TPDINPAMXXKRGIRYGR 19
 Db 263 TPDITKAWAIRHVRIND 261
 RESULT 8
 ID YAAI_ECOLI STANDARD; PRT; 476 AA.
 AC p30143;
 DT 01-JUN-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative transporter yaaI.
 GN YAAI OR B0007.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562;
 RX MEDLINE=92334977; PubMed=1610901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RA Isono K., Mizobuchi K., Nakata A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 0-2.4 min region."
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick N.A., Goeden W.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SODIUM/ALANINE SYMPORTER FAMILY
 (SAP). STRONG, TO H. INFLUENZAE H10183.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: D10483; -; NOT ANNOTATED_CDS.
 DR EMBL: AE000111; AAC73118.1; -;
 DR Ecocore: E01155; YaaI.
 DR InterPro: IPR002293; Aa/rel_prtaseel.
 DR InterPro: IPR001463; Na/Ala_symprt.
 DR Pfam: PF01235; Na_Ala_sympr.1.
 DR PRINTS: PRO0175; NAALASMPORF.
 DR TIGRFA: TIGR00835; agcs.1.
 DR PROSITE: PS00673; NA_ALANINE_SYM.1.
 KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
 KW Symport; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.

FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 300 320 POTENTIAL.
 FT TRANSMEM 351 371 POTENTIAL.
 FT TRANSMEM 391 411 POTENTIAL.
 FT TRANSMEM 414 434 POTENTIAL.
 SQ SEQUENCE 476 AA; 51662 MW; 2FEB2E12E126E63 CRC64;
 Query Match 39.4%; Score 43; DB 1; Length 476;
 Best Local Similarity 44.4%; Pred. No. 11;
 Matches 8; Conservative 2; Mismatches 2; Indels 6; Gaps 1;
 QY 3 DIN-----PAMXXKGI 14
 Db 120 DYNGEFGGPRGYMARGL 137
 RESULT 9
 ID NRPI_YEAST STANDARD; PRT; 719 AA.
 AC P32770; Q12228;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Asparagine-rich protein (ARP protein).
 GN NRPI OR ARP1 OR ARP OR YDI167C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 NCBI_TaxID=4932;
 RX MEDLINE=93247548; PubMed=8483449;
 RA Wehner E.P., Rao E., Brendel M.;
 RT "Molecular structure and genetic regulation of SFA, a gene
 responsible for resistance to formaldehyde in Saccharomycos
 cerevisiae, and characterization of its protein product."
 RL Mol. Gen. Genet. 237:351-358(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AH22;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick N.A., Goeden W.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (Potential).
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: X68020; CAA48159.1; -;
 DR EMBL: Z67750; CAA91579.1; -;
 DR EMBL: Z74215; CAA98741.1; -;
 DR PIR: S31139; S31139.
 DR SCD: S0002326; NRPI.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR InterPro: IPR001876; Znf_RangBP.
 DR Pfam: PF00076; rrm.1.
 DR Pfam: PF00641; zf_RanBP.2.
 DR SMART: SM00360; RRM.1.
 DR SMART: SM00547; Znf_RBP.2.
 DR PROSITE: PS00102; RRM.1.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE_NRG.
 DR PROSITE: PS01358; ZF_RANBP2.1; 2.
 DR PROSITE: PS01099; ZF_RANBP2.2; 2.
 KW Nuclear protein; Zinc-finger; RNA-binding; Repeat.

GN M0762.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=868087;
 RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Koese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii."
 RL Science 273:1058-1073(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: STRONG, TO M.JANNASCHII M0762 AND TO S.POMBE MALATE
 CC PERMEASE (MAE1).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U67521; AAB98753.1;
 DR TIGR: M0762;
 DR InterPro: IPR004695; C4dic_mal_transp.
 DR Pfam: PF03595; C4dic_mal_tran; 1.
 DR TIGRfams: TIGR00816; tcd; 1.
 KW Hypothetical protein; Transmembrane; Transport; Complete proteome.
 FT TRANSMEM 8 28 POTENTIAL.
 FT TRANSMEM 39 59 POTENTIAL.
 FT TRANSMEM 79 99 POTENTIAL.
 FT TRANSMEM 108 128 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 276 296 POTENTIAL.
 FT TRANSMEM 304 324 POTENTIAL.
 SQ SEQUENCE 342 AA; 39534 MW; 08EECE3EC4955D8 CRC64;
 Query Match 37.6%; Score 41; DB 1; Length 342;
 Best Local Similarity 53.3%; Pred. No. 17;
 Matches 8; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
 QY 4 INPAMYXXRGIRPVG 18
 Db 139 VNPGRY----IPPGV 149
 ID Y576.METUA STANDARD; PRT; 347 AA.
 AC 057996;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M0762.
 GN M0762.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Testis.
 RX MEDLINE=88318927; PubMed=2901039;
 RA Singh S., Lowe D.G., Thorpe D.S., Rodriguez R., Kuang W.-J.,
 RA Dangott L.J., Chinkers M., Goeddel D.V., Garbers D.L.;
 RT "Membrane guanylate cyclase is a cell-surface receptor with homology
 to protein kinases."

RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=868087;
 RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Koese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii."
 RL Science 273:1058-1073(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: STRONG, TO M.JANNASCHII M0762 AND TO S.POMBE MALATE
 CC PERMEASE (MAE1).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U67506; AAB98567.1;
 DR TIGR: M0576;
 DR InterPro: IPR004695; C4dic_mal_transp.
 DR Pfam: PF03595; C4dic_mal_tran; 1.
 DR TIGRfams: TIGR00816; tcd; 1.
 KW Hypothetical protein; Transmembrane; Transport; Complete proteome.
 FT TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 46 66 POTENTIAL.
 FT TRANSMEM 84 104 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 249 269 POTENTIAL.
 FT TRANSMEM 283 303 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 SQ SEQUENCE 347 AA; 39556 MW; 632F7671A31DE183 CRC64;

Query Match 37.6%; Score 41; DB 1; Length 347;
 Best Local Similarity 53.3%; Pred. No. 17;
 Matches 8; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
 QY 4 INPAMYXXRGIRPVG 18
 Db 146 VNPGRY----IPPGV 156
 ID CYGR_ARBPV STANDARD; PRT; 986 AA.
 AC P11528;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Resact receptor precursor (Guanylate cyclase) (EC 4.6.1.2).
 OS Archaea punctulata (Punctate sea urchin).
 CC Echinoidae; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 CC Euryarchaeota; Echinoidae; Echinacea; Arbacoidae; Arbacidae; Arbacia.
 OX NCBI_TaxID=7641;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Testis.
 RX MEDLINE=88318927; PubMed=2901039;
 RA Singh S., Lowe D.G., Thorpe D.S., Rodriguez R., Kuang W.-J.,
 RA Dangott L.J., Chinkers M., Goeddel D.V., Garbers D.L.;
 RT "Membrane guanylate cyclase is a cell-surface receptor with homology
 to protein kinases."

Thu May 1 15:54:11 2003

seq74-plus-73.rsp

SEQ ID NO: 74 fused to SEQ ID NO: 73
Database: SwissProt-40
AC NO: P81278

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:36:27 ; Search time 11 Seconds
(without alignments)
124.429 Million cell updates/sec

Title: SEQ74-PLUS-73

Perfect score: 156
Sequence: 1 SRXHXSHKRPDPINPAMYXKRGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the total score distribution,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147	94.2	83	PRRP_RAL	P81278 rattus norv
2	147	94.2	87	PRRP_HUMAN	P81277 homo sapien
3	146	93.6	98	PRRP_BOVIN	P81264 bos taurus
4	65	41.7	4486	DYN9_HUMAN	O94yc9 homo sapien
5	62	39.7	2476	ZAN_PIG	P20471 rhizobium m
6	62	39.7	2832	NDVB_RHIME	O57408 haemophilus
7	61	39.1	999	HGP4_HAELN	O94c10 homo sapien
8	61	39.1	1997	OTOF_MOUSE	O94ef1 mus musculu
9	61	39.1	1997	OTOF_MOUSE	P34082 drosophila
10	60	38.5	873	FAS2_DROME	O26543 methanobact
11	60	38.5	962	UVRA_METTA	P05200 mus musculu
12	60	38.5	3866	HRX_MOUSE	O03164 homo sapien
13	60	38.5	3969	HRX_HUMAN	P98160 homo sapien
14	60	38.5	4393	PGHM_HUMAN	O30565 chlamydomon
15	60	38.5	4568	DYHB_CHIRE	O88854 mus musculu
16	59	37.8	371	GALS_MOUSE	O08726 rattus norv
17	59	37.8	372	GALS_RAT	O08725 rattus norv
18	59	37.8	637	MTHA_DROME	O84509 escherichia
19	59	37.8	940	UVRA_ECO57	P07671 escherichia
20	59	37.8	940	UVRA_ECO57	P31734 salmonella
21	59	37.8	941	UVRA_SATIU	O03745 bacillus th
22	59	37.8	1138	C7AB_BACTU	O03746 bacillus th
23	59	37.8	1138	C7AB_BACTU	O45779 bacillus th
24	59	37.8	1555	POLE_MOUSE	P35574 o glycoen
25	59	37.8	3341	POLG_MOUSE	P35575 m genome po
26	59	37.8	4466	DYHC_TRIGR	P23098 tripunctatus
27	59	37.8	6486	TYCC_BACCH	O30409 b tyrocidin
28	58	37.2	814	POL_IPMAL	P12894 mouse intra
29	58	37.2	940	UVRA_YIBEA	O94kms vldrio choi
30	58	37.2	960	UVRA_TREPA	O83527 treponema p
31	58	37.2	2037	FAST_CANAL	P34731 c fatty aci
32	58	37.2	2225	PYRI_MESAU	P08955 mesocricetu
33	58	37.2	2411	DAB_DROME	P98081 drosophila

34	58	37.2	2715	1	TRX2_HUMAN	O9um6 homo sapien
35	58	37.2	4590	1	FATH_HUMAN	Q14517 homo sapien
36	57	36.5	387	1	GALS_HUMAN	O43603 homo sapien
37	57	36.5	491	1	ADRO_HUMAN	P22570 homo sapien
38	57	36.5	867	1	POL_IPMA	P11368 mouse intra
39	57	36.5	905	1	NIA_FUSOX	P39863 fusarium ox
40	57	36.5	926	1	UVRA_AQUAE	O66911 aquifex aeo
41	57	36.5	986	1	CYGR_ARBPV	P11528 arabacia pun
42	57	36.5	1073	1	YEP4_YEAST	P43564 saccharomyc
43	57	36.5	1165	1	POL_GALV	P21414 gibbon ape
44	57	36.5	1532	1	GDE_HUMAN	P35573 h glycoen
45	57	36.5	1696	1	PKS5_BRACT	O9n15 branchiosto

ALIGNMENTS

RESULT 1
PRRP_RAT STANDARD, PRT, 83 AA.
AC P81278;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (Prp) (Proactin-releasing
DE hormone) [Contains: Proactin-releasing peptide PRRP31; Proactin-
DE releasing peptide PRRP20].
GN PRH.
OS Rattus norvegicus (Rat).
CC Euteleostomi; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-98268781; PubMed-9607765;
RX MEDLINE-98268781; PubMed-9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A proactin-releasing peptide in the brain.";
RL Nature 393:272-276(1996).
[2]
RP TISSUE SPECIFICITY.
RX MEDLINE-99426552; PubMed-10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S.,
RA Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H.,
RA Sunino Y., Fujino M.;
RT "Tissue distribution of proactin-releasing peptide (Prp) and its
RT receptor.";
RL Regul. Pept. 83:1-10(1999).
CC - FUNCTION: Stimulates proactin (PRU) release and regulates the
CC expression of proactin through its receptor GPR10. May stimulate
CC lactotrophs directly to secrete PRU.
CC - TISSUE SPECIFICITY: Widely expressed, with highest levels in
CC medulla oblongata and hypothalamus.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, AB015418; BAA29026.1;
CC Hormone; Amlation; Signal; Cleavage on pair of basic residues.
CC BY SIMILARITY.
FT SIGNAL. 1 21
FT PEPTIDE 22 52 PROACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 52 PROACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 52 52 AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ SEQUENCE 83 AA; 9215 MW; DQC75A264EE4F29 CRC64;
Query Match 94.2%; Score 147; DB 1; Length 83;

Best Local Similarity 83.9%; Pred. NO. 4.5e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMVYXRGIRPVGRF 31
DB 22 SRAHSHMETRTPDINPAMVYXRGIRPVGRF 52

RESULT 2

PRRP_HUMAN

ID PRRP_HUMAN STANDARD; PRT; 87 AA.

AC P81277;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PRRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRR31; Proactin-releasing peptide PRR20].
GN PRR.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]

SEQUENCE FROM N.A.

TISSUE=Brain;

MEDLINE=98268781; PubMed=9607765;

RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;

"A prolactin-releasing peptide in the brain."
Nature 393:272-276(1998).
[2]

TISSUE SPECIFICITY.

MEDLINE=9942652; PubMed=10498338;

RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H., Sumino Y., Fujino M.;

"Tissue distribution of prolactin-releasing peptide (PRRP) and its receptor."
Regul. Pept. 83:1-10(1999).

-1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.

-1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB015419; BAA29027.1;
DR MIM: 602663;
KW Hormone; Amidation; Signal.
FT SIGNAL 1
FT PEPTIDE 23 53
FT PEPTIDE 34 53
FT MOD RES 53 53
SQ SEQUENCE 87 AA; 9659 MW; 229A2F3F50CF981B CRC64;

Query Match 94.2%; Score 147; DB 1; Length 87;
Best Local Similarity 83.9%; Pred. No. 4.9e-10;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMVYXRGIRPVGRF 31
DB 23 SRAHSHMETRTPDINPAMVYXRGIRPVGRF 53

RESULT 3

PRRP_BOVIN
ID PRRP_BOVIN STANDARD; PRT; 98 AA.

AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PRRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRR31; Proactin-releasing peptide PRR20].
GN PRR.

OS Bos taurus (Bovinae).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
TISSUE=Brain;
MEDLINE=98268781; PubMed=9607765;

RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;

"A prolactin-releasing peptide in the brain."
Nature 393:272-276(1998).

-1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.

-1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB015417; BAA29025.1;
DR Hormone; Amidation; Signal; Cleavage on pair of basic residues.
KW Hormone; Amidation; Signal.

FT SIGNAL 1 22
FT PEPTIDE 23 53
FT PEPTIDE 33 53
FT MOD RES 53 53
SQ SEQUENCE 98 AA; 10544 MW; 08AC35A13B0FA908 CRC64;

Query Match 93.6%; Score 146; DB 1; Length 98;
Best Local Similarity 83.9%; Pred. No. 7.8e-10;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMVYXRGIRPVGRF 31
DB 23 SRAHSHMETRTPDINPAMVYXRGIRPVGRF 53

RESULT 4
ID PRRP_HUMAN STANDARD; PRT; 4486 AA.

AC Q9NYC9; OSNO28; O95494;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ciliary dynein heavy chain (Axonemal dynein heavy chain) (dynein heavy chain 9).
GN DNAB9 OR DNAB17L OR DNAB1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]

SEQUENCE FROM N.A.
RA Reed W., Moats-Staats B.M., Carson J.L., Leigh M.W., Collier A.M.;
RT "A ciliary dynein heavy chain whose expression is upregulated in differentiating airway epithelium."

RL Submitted (AFR-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Nasal Epithelium;
 RA Bartoloni L., Blohin J.L., Maiti A., Sainsbury A., Rossier C.,
 RA Gehrig C., She J.X., Marron M.P., Lander E., Meeks M., Chung E.,
 RA Jorissen M., Scott H.S., Detozler Blanchet C.D., Gardiner M.,
 RA Antonarakis S.E.;
 RT "Axonemal beta heavy chain dynein DNAB9: cDNA sequence, genomic
 RT structure and investigation of its role in primary ciliary
 RT dyskinesia.";
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1874-1974 FROM N.A.
 RC TISSUE-Nasal Polyps;
 RA Maiti A.K., Maitel M.G., Jorissen M., Voiz A., Ziegler A.,
 RA Bouvagnet P.;
 RT "Chromosomal localization of human dynein heavy chain genes";
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE-20558134; PubMed-11104725;
 RA Reed W., Carson J.L., Moats-Staats B.M., Lucier T., Hu P.C.,
 RA Brighton L., Gambing T.M., Huang C.H., Leigh M.W., Collier A.M.;
 RT "Characterization of an axonemal dynein heavy chain expressed early in
 RT alveary epithelial cillogenesis.";
 RL AM. J. Respir. Cell Mol. Biol. 23:734-741(2000).
 CC -1- FUNCTION: FORCE GENERATING PROTEIN OF RESPIRATORY CILIA. PRODUCES
 CC FORCE TOWARDS THE MINDS ENDS OF MICROTUBULES. DYNEIN HAS ATPASE
 CC ACTIVITY.
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
 CC INTERMEDIATE AND LIGHT CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF257737; AAF69004.1; -
 CC EMBL: AJ404468; CAB94756.1; -
 CC EMBL: AJ133088; CA10561.1; -
 CC Gene: HGNC:2953; DNAB9.
 CC MIM: 603330; -
 CC
 CC DR InterPro: IPR004273; Dynein_heavy.
 CC DR Pfam: PF03028; Dynein_heavy; 1.
 CC KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
 CC FT DOMAIN 381 410 COILED COIL (POTENTIAL).
 CC FT DOMAIN 504 529 COILED COIL (POTENTIAL).
 CC FT DOMAIN 639 662 COILED COIL (POTENTIAL).
 CC FT DOMAIN 752 823 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1326 1355 COILED COIL (POTENTIAL).
 CC FT DOMAIN 3051 3154 COILED COIL (POTENTIAL).
 CC FT DOMAIN 3285 3341 COILED COIL (POTENTIAL).
 CC FT DOMAIN 3640 3675 COILED COIL (POTENTIAL).
 CC FT NP_BIND 1870 1877 ATP (POTENTIAL).
 CC FT NP_BIND 2151 2158 ATP (POTENTIAL).
 CC FT NP_BIND 2478 2485 ATP (POTENTIAL).
 CC FT NP_BIND 2825 2832 ATP (POTENTIAL).
 CC FT CONFLICT 2505 2505 L -> V (IN REF. 2).
 CC FT CONFLICT 3578 3578 T -> A (IN REF. 2).
 CC FT CONFLICT 4374 4374 I -> M (IN REF. 2).
 CC SQ SEQUENCE 4486 AA; 511927 MW; 996BDFEBOH3BE1 CRC64;
 CC
 CC Query Match 41.7%; Score 65; DB 1; Length 4486;
 CC Best Local Similarity 0.6%; Pred. No. 5.4e+02;
 CC Matches 19; Conservative 1; Mismatches 10; Indels 3343; Gaps 5;
 CC
 CC 2 RXHXSN----- 8

Db 163 RRAHSIQCDLSTVILEQVKGKTLPLPAGSEKMEPADSKSETVIDSKSVIAIESAVI 222
 QY 9 ----- 8
 Db 222 KMSQVQVYVVKNESSQPLLQGENPTPKVELEFWKSRVEDLKYINQLTIVVNGMAKILD 282
 QY 9 ----- 8
 Db 283 KLOSSFPAFKMYADVVAAAEADIVHVLPIQRHLEALENAEFEPVKQPLRLHHV 342
 QY 9 ----- 8
 Db 343 CLTATCKSYKSPGRILYLDQETCNLLIQASNTLSPEDLIRSEVESQRLQVSDTLS 402
 QY 9 ----- 8
 Db 403 FPKQFQDRENTLHYFEKQNEVKEMDQSSLVFRLDGLIGRLHYVGLKTLADPHKL 462
 QY 9 ----- 8
 Db 463 GKVESSVGRNALSQVQVQMEEPQENTRLSGSSDCLYLQSTDFENDVSEFNQVEDL 522
 QY 9 ----- 8
 Db 523 DRLGTITLQAFDAPGLEHAFKLIDINGNLLERPLVARDTSKTYLYIQNFNKDLAVR 582
 QY 9 ----- 8
 Db 583 MYSQHQEAEALGPSYKKNMPTVAGLFWNQELRQIQGPFNSPGRITHPCKESABG 642
 QY 9 ----- 8
 Db 643 RMOQTEEDLSLEKETRLYEDWCRTVSEKSYNLSQLKRDPEKEITINENPOLIS 702
 QY 9 ----- 8
 Db 703 VLKESYLEPRMKRMPTAAANSSRDFYQLYANLEMANVTKYKTLVEFPLVE 762
 QY 9 ----- 8
 Db 763 EELQNDILRLAEBELNMTKEGICDYTEITSIHDEQRLQTKDNVEIQNMKTW 822
 QY 9 ----- 8
 Db 823 TPIFTKDKRESLISLDRDRDREKYYNLKESGLIKHALVQENTLGSADPTSNITKT 882
 QY 9 ----- 8
 Db 883 YVNSIDNLLNGFLATICSILKYLENTECKAGLTFPEAQSLAIPBLVYPSLSGVK 942
 QY 9 ----- 8
 Db 943 GGPCEIVGILTSIPRISLVRPLSPQNSPHQVLDGIPDLANMRRLMERVQRMGL 1002
 QY 9 ----- 8
 Db 1003 CCGYSTSQSYLYVEDRKVYLGQFLLYGHLPEPEIEHVEDGIDENPELLSQFVQI 1062
 QY 9 ----- 8
 Db 1063 DSVETLYEVCRLPIKVFQGMKIDIRPKASILNITKRSLEFKQHLVDVTHSLANL 1122
 QY 9 ----- 8
 Db 1123 DAFIKSESGLIKVKEKDFQGLVEINQHLNAVYRQSNIDEMEPILKQTEILKTYEQE 1182
 QY 9 ----- 8
 Db 1183 LPETVFKQLEELPERKNNIKVATITVQVAPLQANEYTLIRQCTAFDAEQQFWDFH 1242
 QY 9 ----- 14
 Db 1243 KEAPRFPSIHPHQMLDARHIEIQMESTMASISBSALFEVNPVDYKQILNQCKEVCQL 1302


```

QY 15 ----- 14
Db 1303 KEMLDIGWYSSIHAWETTPMKNINVEAMELECKOFARHIRNLDKVRAMDAFTGLEST 1362
QY 15 ----- 14
Db 1363 VMTLSSLRAVAELONPAIBERHFRQLOMGATGVSTMDODPTLAHLIQLQHLHYEDEVG 1422
QY 15 ----- 14
Db 1423 IVDKANEKMEKTELKELQTTWAGMEFQYEPHPTNPPLCSDEDLLEVLBNQVOLONT 1482
QY 15 ----- 14
Db 1483 VMSKYVAFLEEVSGWOKLSTVDAVISIMFVQRTWHLSEJFSGSEDJRAQLPODSKR 1542
QY 15 ----- 14
Db 1543 FEGIDIDFKELAYDAQKIPNVVQTPNKGLEYEKLEDOGRGLCEKALAEYLDTKRLAFP 1602
QY 15 ----- 14
Db 1603 RFTFLSSDLIDJLSNGTAPQOVORHLSKLEFDMNAKRFQOLDSGEPTKSLGNYSKEEB 1662
QY 15 ----- 14
Db 1663 YVAFSEPODCSGOYEIWLNLHVLAGMKAVVHHEMTBGVTAEEKPREQWLFDPNAQVALYC 1722
QY 15 ----- 14
Db 1723 TQIWMTEVGNAFARLEEGYESAKDYKQVQAKTLITMLJGQSKRGROKIMPTCTI 1782
QY 15 ----- 14
Db 1783 DVHARDVYAKKIAQKVNQAQFLMLSQLRNRNDEYKHCANICDAQFLXSYELGNTPR 1842
QY 15 ----- 14
Db 1843 LVTPPLDRCYITLQSLHLTWGAPAGPAGTGKTEFTKDLGRALGILVYENCSEQMOY 1902
QY 15 ----- 14
Db 1903 KSGCNTRYKGLAQGTAMGCFDEENRISVEVLVAVQVKSJODAIRKQWFSPLGEEISL 1962
QY 15 ----- 14
Db 1963 NPSVGIFTMNPYAGRTLEPENLKSIFPCAMVVPDFELCEIMLVAEGFIENQSLARK 2022
QY 15 ----- 14
Db 2023 FTTLYQJLCKELLSKODHYDWGLRAIKSVLVAGSLKRGDEPRDQYLMRSLRDNFIPKI 2082
QY 15 ----- 14
Db 2083 VTDDMPITFMGLIGLFPALDVPRRDRPFELVYRKALVLDLQAEENFVLAKVQLEELLA 2142
QY 15 ----- 14
Db 2143 VRHSVFWVGAGTGSQVLSLHKTQIMKRPRVWTDLNKRAYNDELPGIINPATGEMK 2202
QY 15 ----- 14
Db 2203 DGLFSSIMRELANTHDCPKWILLDCDIDPMWLESNTVMDNKNVLTLASNERILPLNPTM 2262
QY 15 ----- 18
Db 2263 KLFELISNLTAPATVSRAGIILYNPADLGMNPPVSSWTEKREIQTERANLTLPDKYL 2322
QY 19 ----- 18
Db 2323 PTCIDTJLTKRKKIIPPEOSNVMQVCHLLECLITTEDIPADCPKEIYEHYFVPAIINAF 2382

```

```

QY 19 ----- 18
Db 2383 GGAMVQDOLVDYRAESKRWLEFETVAKPPSOGTLPDYIIDEPTFKTEPNKIVPQFEED 2442
QY 19 ----- 18
Db 2443 PEMPLQACLVHSETIRVCYEMERLMAORPVMVLGTAGTGSVLGAKLASLDEPVALY 2502
QY 19 ----- 18
Db 2503 KNLPEYTTSSMLQAVLEKPLEKAGNRYGPPGKKLITPIDDKNMEVDATGVQPH 2562
QY 19 ----- 20
Db 2563 IIRQHLDYGHVYDRSKLSLKEITNVQVYSCMNPAGSFTINPRLQRFNVFVLSPPGADA 2622
QY 21 ----- 20
Db 2623 LSTIYSIILTOHLKLGNPASLOKSIPPLIDLALFHOKIATFTLPTGIRKHYIFNLRF 2682
QY 21 ----- 20
Db 2683 ANIFQGLFSSVECYKSTWDJIRLTLHESNRVYRDKMVEKDFDLFDKIQTEVLKATPDD 2742
QY 21 ----- 20
Db 2743 IEDVPEQTSPLNYCHFANGIGEPKMPYQSWELLTQTLVEALEENHNEVNTVMDLVLEED 2802
QY 21 ----- 20
Db 2803 ANRHVCNINRILLESPPGNALLVGVGSGKOSLITRLAFTISMDVFOJTLRGYQJODEKM 2862
QY 21 ----- 20
Db 2863 DLASLCRAGVKNLWVPLATDAQVADERFTVLIJDLASGEIPDLYSDDDEVNTISVR 2922
QY 21 ----- 20
Db 2923 NEVKSGLVDNRENGCKFEIDIRIRQLKVTLCFSPVGNLRYVSRKFPALVNCIAIHWFH 2982
QY 21 ----- 28
Db 2983 EMPQALSESUSLRFQNTQEGIEPTVKQISKEPAFVHTSVNOTSQSYLSNEORRYTTPK 3042
QY 29 ----- 28
Db 3043 SPLEFIRLYQSLHHRKRELKCTERLENGILKXHSQAQVDDLKAKIQAQVELKQKNE 3102
QY 29 ----- 28
Db 3103 DADKILQVNVGETDKYSREKMADEEOKVAVIMLEVKOKDCBEDLAKAEPALTAQA 3162
QY 29 ----- 28
Db 3163 ALNLTNKTMLTELKSEFSPPLAVSVSAVMTAPRGRVPRDNRKAKAYTAKVDSFL 3222
QY 29 ----- 28
Db 3223 DSIINPNKENIHENGLKAIKRPYLQDPERNPEFVATKSYAAAGLCSVNTIYREFEVPDV 3282
QY 29 ----- 28
Db 3283 EPRQALNKAJADLTAOQKLAIRAKIAHLNENLAKILAREKATADLQOQAEAYTA 3342
QY 29 ----- 28
Db 3343 VTIISLANRLVGLASENVRMADAVONFQOERTLCGDILLTAPISYIGFTTKYRQSL 3402
QY 29 ----- 28
Db 3403 DRTWRPYSQLEKPIPYTPALDPLKMLMDADVAAWONEGLPADMSVEHATILINCERN 3462
QY 29 ----- 28

```


DB 3463 PLMBP01OGIKNKIKGEDIKRTYQIOCKGVLITTEQALEAGAVLIEESIDPYLG 3522
 OY 29 -----GRF 31
 DB 3523 PLLGKREVIKGRF 3535
 RESULT 5
 ZAN_PIG STANDARD: PRT: 2476 AA.
 ID ZAN_PIG Q28983;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zonadhesin precursor.
 GN ZAN.
 OS Sus scrofa (Pig).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 CC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 823-830; 859-872; 883-890;
 RP 920-925; 960-967; 1235-1244; 1349-1354; 1518-1532; 1624-1656;
 RP 1658-1667; 1777-1795 AND 1914-1921.
 RC STRAIN=Meishan; TISSUE=Testis;
 RX MEDLINE=96064658; PubMed=7592795;
 RA Hardy D.M., Gathers D.L.;
 RT "A sperm membrane protein that binds in a species-specific manner to
 RT the egg extracellular matrix is homologous to von Willebrand
 RT factor".
 RT J. Biol. Chem. 270:26025-26028(1995).
 RT J. Biol. Chem. 270:26025-26028(1995).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: IN TESTIS, PRIMARILY IN HAPLOID SPERMATIDS.
 CC NOT IN LUNG, LIVER, HEART, SPLEEN, BRAIN, KIDNEY, EPIDIDYMIS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE WVD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- FROM THE ZONADHESIN THAT BINDS TO THE EGG EXTRACELLULAR MATRIX.
 CC PROCESSING MIGHT OCCUR DURING SPERM MATURATION AND/OR
 CC CAPACITATION.
 CC -1- SIMILARITY: CONTAINS 2 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4.5 WVD DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: UA0024; AAC48486.1; -.
 DR HSSP: P56682; IGV.
 DR InterPro: IPR000561; EGF-1like.
 DR InterPro: IPR000998; MAM_domain.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR003338; TIL_Cysrich.
 DR InterPro: IPR001007; WVF_C.
 DR InterPro: IPR001846; WVF_D.
 DR Pfam: PF00094; wvd; 4.
 DR Pfam: PF00629; MAM; 2.

DR Pfam: PF01826; TIL; 5.
 DR Pfam: PF02345; TILA; 5.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00137; MAM; 1.
 DR SMART: SM00214; WVC; 2.
 DR SMART: SM00216; WVD; 4.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS00600; MAM_2; 2.
 KW Signal; glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
 KW Repeat.
 FT SIGNAL 1 29
 FT CHAIN 30 2476
 FT DOMAIN 30 2418
 FT TRANSMEM 2419 2439
 FT DOMAIN 2440 2476
 FT DOMAIN 31 144
 FT DOMAIN 147 312
 FT DOMAIN 319 687
 FT DOMAIN 688 739
 FT DOMAIN 800 1184
 FT DOMAIN 1185 1573
 FT DOMAIN 1574 1968
 FT DOMAIN 1969 2370
 FT DOMAIN 2366 2402
 FT DISULFID 2370 2381
 FT DISULFID 2375 2380
 FT DISULFID 2382 2401
 FT CARBOHYD 109 109
 FT CARBOHYD 269 269
 FT CARBOHYD 735 735
 FT CARBOHYD 758 758
 FT CARBOHYD 833 833
 FT CARBOHYD 1134 1134
 FT CARBOHYD 1339 1339
 FT CARBOHYD 1448 1448
 FT CARBOHYD 1544 1544
 FT CARBOHYD 1596 1596
 FT CARBOHYD 1654 1654
 FT CARBOHYD 1843 1843
 FT CARBOHYD 1965 1965
 FT CARBOHYD 2122 2122
 FT CARBOHYD 2165 2165
 FT CARBOHYD 2178 2178
 FT CARBOHYD 2329 2329
 FT CARBOHYD 2359 2359
 FT CONFLICT 823 823
 FT CONFLICT 923 923
 FT CONFLICT 965 965
 FT CONFLICT 1241 1241
 SQ SEQUENCE 2476 AA; 270364 MW; A13B690375A6548C CRC64;
 Query Match 39.7%; Score 62; DB 1; Length 2476;
 Best Local Similarity 1.6%; Pred. No. 4e+02;
 Matches 19; Conservative 2; Mismatches 7; Indels 1133; Gaps 5;
 OY 4 HXH-----SMEKRT-----
 DB 762 HCRCPSSMRECOITRCGHTVYQIOCKGVLITTEQALEAGAVLIEESIDPYLG 821
 OY 13 -----
 DB 822 KCTYILAPRCNLTETFRVLYKKEEROEGVSCSKYTYVTLPESTYTLKGRHLYVGQ 881
 OY 13 -----
 DB 882 RYTLPAIPRGVFLAPSGRFVELQAFGLRYWMDQDLFWVSPSTFSKGLGLGVDYD 941
 OY 13 -----
 DB 942 DSSNDKQKDGSPADADEKELSSWQTSSEDAQQCEENQVSPSCNTALQNTWSPGPFCCG 1001


```

QY 12 ----- 11
Db 382 KQRLAEKRTGSPSTFQHLINSVKRDMFALAGPNILLITLAMIYVAFVSPMDIPSCA 441
QY 12 ----- 11
Db 442 KLIMLLPALPASSEGAMGLENVTLLFAKPSRLVGEERLIDPEDARKLVVPCILAKRD 501
QY 12 ----- 11
Db 502 HVDLVRNLEHYLANPGEIYFALLSDMADSKSEAPADVLEAKHEIASLARYAY 561
QY 12 ----- 11
Db 562 DKRTRETLHRRRLYNAGVWGRKRGKLEHLLNLLRGDRDTSFLOGANWEGVQY 621
QY 12 ----- 18
Db 622 VMTLSDFTLMDATYKLVGKLYHPTLNRPVNPRTQEVVTVGSLQPRVTSPLTGSSEAS 681
QY 19 ----- 18
Db 682 AFQRLFTINRGIDPYVFTVSDVYQDLAGEGFTGKLYHDAFPAALSKRIENAVLSND 741
QY 19 ----- 18
Db 742 LLEGYARCALVTDIELVEDPPIREVEMSRONRRARGMDLLPYTFNPKGLSMIGRK 801
QY 19 ----- 18
Db 802 MYDNLRSLLIPAVMLAASVWGYMEPTPALIQVLIIFSEVAPTLISGIMPRNDI 861
QY 19 ----- 18
Db 862 VARAHLETVLSDIRAAHQVALRIVEIAHNAAMADAIYRSLYRFEVSRKLEWRTAAQ 921
QY 19 ----- 18
Db 922 VOSAGHGTGDTFRAMTAPALALVSLALAISTDGLPGLPALIWAASPAVAVTSQ 981
QY 19 ----- 18
Db 982 SATEDQIVVSEAEIEMKRIARTWRYEAFVTAQNLPPDNQETPPVLAERTSPT 1041
QY 19 ----- 27
Db 1042 NIGVLLSVASARSIGWIGFEETITRLAQTLATIDRMKRYNGHLENNYRTNGLIEPMBRY 1101
QY 28 ----- 27
Db 1102 VASVDSGNLAGNLIAVSSKCREMAEAPSAHVQGNLDIGDAVAILKEALINELPDDKTVR 1161
QY 28 ----- 27
Db 1162 PLRLVEERINAGFQNALAAVKKERELASTRVINLAVLRDMKLTVNDHEVRYQSGEV 1221
QY 28 ----- 27
Db 1222 ATWAGSLVAAACEAHADGVFDGAIEALRQRLVLEKARADTAQMDSEFLRPERRLIS 1281
QY 28 ----- 27
Db 1282 IGRVYANNELEACTDCLASERLTSLAIAKGLDPTHEWYKLGRIIVFIGARGALVSMS 1341
QY 28 ----- 27
Db 1342 GSNFETLMPPLVMOERGSLINQTNLVVQEGINHGRLCTGKTSAAFNARDELITYQ 1401
QY 28 ----- 27
Db 1402 YNFEVFTLGLKRGIGQNAVTAPYASIIACMKDPSALANLRLREVGALGAYGNDAVD 1461
QY 28 ----- 27

```

```

Db 1462 FTPTREPGQCAVVRNRYANHNGNSVAAVANVENGQLEWFRADPVIEAELLIOEKA 1521
QY 28 ----- 27
Db 1522 PRDIPVAAKREPALGKQADLRLPEVRVVEDEPINODRETVILSNCHYSVMLTATGAGY 1581
QY 28 ----- 27
Db 1582 ARMGQSVTRTPDPVEDRIGTIFILNDIYVGDWMSATAPBRAPGEKTVTRGDOKAEF 1641
QY 28 ----- 27
Db 1642 VKTVGDLTSEVECIYAEHDAEGRRVILLNTGDRFIPVTSAPVLAAMDADSSHPTE 1701
QY 28 ----- 27
Db 1702 SKMFLRELSHGCVIYWSRNKRSFGDDTIEVALYTDNAGSERHIOAETDRRRFLGOCR 1761
QY 28 ----- 27
Db 1762 TLAEAAFPDGTATLSTGDTFTDPTVSLRVRVPACKVSVIFWIIAADRGCVDAID 1821
QY 28 ----- 31
Db 1822 RYRREPEIFNHELIAWTRSOYQMRHVIGITSKEAASQMLGRY 1863

```

RESULT 7

```

HGPI_HAETIN STANDARD; PRT; 999 AA.
ID HGPI_HAETIN 057408; P96344; 086244;
AC 057408; P96344; 086244;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable hemoglobin and hemoglobin binding protein 4
DE precursor.
GN H11565/H11567.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907,
RZ MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerkovitz A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Nanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerkovitz A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.,
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP CONCEPTUAL TRANSLATION.
RA Couderc E.;
RL Unpublished observations (AUG-2001).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR TNE
CC HEMOGLOBIN/HAPToglobulin COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Outer membrane (by similarity).
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA

```

REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISMATCHING. ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY). -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY; HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY. -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE INTRODUCED IN POSITION 289 AND IN THE REPEATS REGION TO MAXIMIZE THE SIMILARITY WITH OTHER ORTHOLOGS.

 THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/> OR SEND AN EMAIL TO license@isb-sib.ch).

CC EMBL, U32831; AAC23213.1; ALT_SEQ.
 CC EMBL, U32831; AAC23214.1; ALT_SEQ.
 CC TIGR, H1565; -
 CC TIGR, H1567; -
 CC InterPro: IPR000531; TONB_BOX.
 CC Pfam, PF00593; TONB_BOX; 1.
 CC PROSITE, PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 CC PROSITE, PS01156; TONB_DEPENDENT_REC_2; 1.
 CC Outer membrane; Transport; TONB box; Multigene family; Signal;
 CC Receptor; Repeat; Complete proteome.
 CC SIGNAL 25 24 POTENTIAL.
 CC CBAIN 25 999
 CC
 CC DOMAIN 26 49 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
 CC REPEAT 26 29 HAPTOGLOBIN BINDING PROTEIN 4.
 CC REPEAT 30 33 6 X 4 AA TANDEM REPEATS OF P-T-H-Q.
 CC REPEAT 34 37 2.
 CC REPEAT 38 41 3.
 CC REPEAT 42 45 4.
 CC REPEAT 46 49 5.
 CC REPEAT 58 65 6.
 CC SITE 58 65 TONB_BOX.
 CC SITE 982 999 TONB_C-TERMINAL_BOX.
 CC SEQUENCE 999 AA; 116314 MF; DAFCD4EB7000A876 CRC64;

Query Match 39.1%; Score 61; DB 1; Length 999;
 Best Local Similarity 2.2%; Pred. Hc. 99;
 Matches 16; Conservative 4; Mismatches 11; Indels 713; Gaps 4;

OY 1 SRXHSSE-----9
 DB 241 SKRHGHELENDYKNGNDIQEKERKADPYITRESLVKESFSPTEHREPTVASDYLQ 300
 OY 10-----9
 DB 301 NSRGHDSYNAVKTYIINKDEELRHNDLTKRNVSTYENTYVTFPWDTLKLSYGOQR 360
 OY 10-----15
 DB 361 IYTRARTDYCOGNEKCDYKPNGLALKEGKYVDKNGDPVELKLVDEDEGQKRRHGVAK 420
 OY 16-----15
 DB 421 YNNPVSASGTNDALFVKQLSPSEFWLDSIFNCDKPRVRYKYQYSNOPESEKEVELNR 480
 OY 16-----15
 DB 481 TMEINGKFAYESNNRYDRHMLPNKGYLPIDYKERDLNTKQINDLTAKATLFE 540
 OY 16-----20
 DB 541 IENELSYGYAKTKKEMVKAAGYGRNPTWMAERTLGLKSLNGLRCKEDSSYNGLLCP 600
 OY 21-----20

DB 601 RNEPKTSFLIPEVETTKSLYFADNLIKLANMSVDLGYRYDDIKYQPEYIPGTPAIADDM 660
 OY 21-----20
 DB 661 VRELFPPLPPANKDQNGNVTYTPQIRKNAEENIAYINQERFKHSYSLGATPDPLNF 720
 OY 21-----20
 DB 721 LRQVYKSGFRTPTSDLEYTEKHDPFTLPNPNKPREAKNQELATFHHDWGFFSYH 780
 OY 21-----20
 DB 781 VEGTYKQFIDLAYLGSRLNSVSGOARDFOYQYVNVDPRAKYGVINSRLNIGYF 840
 OY 21-----20
 DB 841 FEKLOFNYSYKFTYQGRGLQGRPNALQPKTSYGLGYLDEKDEKDFGADLYTVHSARK 900
 OY 21-----X 21
 DB 901 AKDTYNNFEKQGYKDSAVKRSDDYTLVDYTYIKPVXNVTLQGVYMLDRKYLWES 960
 OY 22 XRGIRPVG-----RF 31
 DB 961 ARSIRKPGTGNLINOGTGAGINRF 984

RESULT 8
 OTOP_HUMAN
 ID OTOP_HUMAN STANDARD; PRT, 1997 AA.
 AC Q9HC10; Q9HC09; Q9Y650; Q9HC08;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, last sequence update)
 DE Otoferrin (Fer-1 like protein, 2).
 GN OTOP OR FERL2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RH [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3), AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=20395831; PubMed=10903124;
 RA Yasunaga S., Grait M., Chardenoux S., Smith T.H., Friedman T.B.,
 RA Lalwani A.K., Wilcox E.R., Petit C.;
 RT "OTOP encodes multiple long and short isoforms: genetic evidence that
 RT the long ones underlie recessive deafness DFNB9.";
 RL Am. J. Hum. Genet. 67:591-600(2000).
 RH [2]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Fetal;
 RX MEDLINE=99206603; PubMed=10192385;
 RA Yasunaga S., Grait N., Cohen-Salmon M., El-Amraoui A., Mustapha M.,
 RA Salem N., El-Kil E., Loiselet J., Petit C.;
 RT "A mutation in OTOP, encoding otoferlin, a FER-1-like protein, causes
 RT DFNB9, a nonsyndromic form of deafness.";
 RL Nat. Genet. 21:363-369(1999).
 CC -1- FUNCTION: Might be involved in the Ca(2+)-triggered synaptic
 CC vesicle-plasma membrane fusion.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (by similarity).
 CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; 1/long (shown here),
 CC 2/short-1, 3/short-2 and 4/short-3; are produced by alternative
 CC splicing.
 CC -1- TISSUE SPECIFICITY: Isoform 1 and isoform 3 are found in adult
 CC brain. Isoform 2 is expressed in the fetus and in adult brain,
 CC heart, placenta, skeletal muscle and kidney.
 CC -1- DISEASE: Defects in OTOP are the cause of nonsyndromic autosomal
 CC recessive deafness 9 (DFNB9).
 CC -1- SIMILARITY: BELONGS TO THE FERLIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 C2 DOMAINS.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on list
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; AF183185; AAC12891.1; -
DR	EMBL; AF183186; AAG12892.1; -
DR	EMBL; AF183187; AAG17468.1; -
DR	EMBL; AF107403; AAD26117.1; -
DR	HSSP; P04410; 1A25.
DR	Genew; HGNC:8515; OTOF.
DR	MIM; 603681; --
DR	MIM; 601071; --
DR	InterPro; IPR000008; C2.
DR	Pfam; PF00168; C2; 8.
DR	SMART; SMO0239; C2; 6.
DR	PROSITE; PS00499; C2 DOMAIN 1; 2.
DR	PROSITE; PS50004; C2 DOMAIN 2; 4.
KW	Transmembrane; Repeat; Alternative splicing; Deafness. CYTOSOLIC (POTENTIAL).
FT	DOMAIN 1 1963
FT	TRANSSEM 1984
FT	DOMAIN 1985 1997
FT	DOMAIN 241 338
FT	DOMAIN 404 514
FT	DOMAIN 947 1052
FT	DOMAIN 1479 1577
FT	DOMAIN 1303 1310
FT	DOMAIN 1314 1320
FT	DOMAIN 1965 1983
FT	VANSPPLIC 1 747
FT	VANSPPLIC 1245 1264
FT	VANSPPLIC 1 690
FT	VANSPPLIC 691 738
FT	
FT	
FT	VANSPPLIC 1943 1997
FT	
FT	
FT	
FT	
FT	
CONFLICT	1088 1088
SEQUENCE	1997 AA; 226735 MM; 39D10CB522063BAE CRC64;

Db	543	TRNTLLDEHODLNEGIGEGVSFARLLGLAVEIDTSPNLTSTEVQDAPFISB	602
QY	20	-----	19
Db	603	CAGKKEEFFLGAFLEASMDIRNGDKPITFEVTIGNYGENVDGLSRPQRPBRKEPGE	662
QY	20	-----	19
Db	663	EEVDLIQNASDDEAGDAGDLASVSTPEPFPQYTDNRNPHLYLEBKPCITYLKSMPDOR	722
QY	20	-----	19
Db	723	RRLYANAMIDHIADKLEBGLNDIQEMIKTEKSYPERLRGVLLELSCGCRFLSLADNQ	782
QY	20	-----	19
Db	783	GHSNTRIDRERLCKMCRLEENGOQARMLRAQVKHTVBDKLRLOQLKRLFLADEP	842
QY	20	-----	19
Db	843	QHSIPDIFTMMSNKKVANAARVPSKDILFIVEERTGMDCAVKTLFLKLGKRGFSA	902
QY	20	-----	19
Db	903	GNTYQAKVELYLMGLSKQKREFLCGLPGSPQEVKAAQSLGLHAFPPSVLYTKKQAPOL	962
QY	20	---YXR-----	23
Db	963	RAHMYQARSLFPAADSGSLDPFARVFFINOSQTEVILNETIACPTWQMLVFNDLFLYGA	1022
QY	24	-----	23
Db	1023	HELKDDPPIVIEIYDODSMGKADFMGRTAKPLVKMADEANCCPFPPOLEYQIYRGN	1082
QY	24	-----GIRP-----	27
Db	1083	ATAGDPLAFLAELLQIGRAGRADLPPIINGPYDVDRGPIIMPYMGIRVLSKYREVILFGL	1142
QY	28	-----	27
Db	1143	RDLKRVNLAQDRBRVDIECAGKQVSSLIHNTKKNPNEVILVKFEVDLPENELLHPL	1202
QY	28	-----	27
Db	1203	NIRVYDCRAFGRYTLVGSHAVSSLRFTYRPPDSAPSMNTYVILLRRCRYLNGGSSSH	1262
QY	28	-----	27
Db	1263	STGEVVTMEBEVYIKLLETWVILDAISEAVYVDVALEEKKKKKKKTAEPEBEEDP	1322
QY	28	-----	27
Db	1323	ESMLDMWSKYFASIDIMKEQLROEPESGIDLEKEBEVDMTEGLKSGMKKEKARAKAEK	1382
QY	28	-----	27
Db	1383	KKKTQSSGSGQSEAPKKPKIDELKVYPKELESEFDNEMLATFNLRLRGKTGDDG	1442
QY	28	-----VGRF 31 	27
Db	1443	STEEERYVGRF 1453	

RESULT 9		
OTOF_MOUSE	OTOF_MOUSE	STANDARD; PRT; 1997 AA.
ID	Q9SEF1: Q9SEF2	
AC	15-JUN-2002 (Rel. 41, Created)	
DT	15-JUN-2002 (Rel. 41, Last sequence update)	
DI	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Otoclerin (Fer-1 like protein 2).	
GN	OTOF OR FERIL2.	

OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OK NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain, and Cochlea;
 RX MEDLINE=20395831; PubMed=10903124;
 RA Yasunaga S., Grati M., Chardenoux S., Smith T.N., Friedman T.B.,
 RA Lalwani A.K., Wilcox E.R., Petit C.;
 RT "OtoF encodes multiple long and short isoforms: genetic evidence that
 RL the long ones underlie recessive deafness DFNB9.";
 RL Am. J. Hum. genet. 67:591-600(2000).
 RN (2)
 RP TISSUE SPECIFICITY.
 RX MEDLINE=99206603; PubMed=10192385;
 RA Yasunaga S., Grati M., Cohen-Salmon M., El-Amraoui A., Mustapha M.,
 RA Salem N., El-Zir E., Loiselet J., Petit C.;
 RT "A mutation in OtoF, encoding otoferlin, a PER-1-like protein, causes
 RL Nat. Genet. 21:363-369(1999).
 CC -1- FUNCTION: Might be involved in the Ca(2+)-triggered synaptic
 CC vesicle-plasma membrane fusion.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (by similarity).
 CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms: 1 (shown here) and 2;
 CC are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Strongly expressed in brain and inner ear. In
 CC the inner ear, it is mainly expressed in the cochlear IHC and
 CC vestibular type I sensory hair cells. Weakly expressed in eye,
 CC heart, skeletal muscle, liver, kidney, lung and testis.
 CC -1- SIMILARITY: BELONGS TO THE PERLIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 C2 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF183183; AAG12989.1; -
 DR EMBL: AF183184; AAG12990.1; -
 DR HSBP: P04410; 1A23.
 DR MGD: MGI:1891247; Ocof.
 DR InterPro: IPR000008; C2.
 DR Pfam: PF00168; C2; 8.
 DR PRINTS: PR00360; C2DOMAIN.
 DR SMART: SM00239; C2; 6.
 DR PROSITE: PS00499; C2_DOMAIN_1; 2.
 DR PROSITE: PS50004; C2_DOMAIN_2; 4.
 KW Transmembrane; Repeat; Alternative splicing.
 FT DOMAIN 1 1963 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1964 1984 POTENTIAL.
 FT DOMAIN 1985 1997 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 240 337 C2 DOMAIN 1.
 FT DOMAIN 403 513 C2 DOMAIN 2.
 FT DOMAIN 946 1051 C2 DOMAIN 3.
 FT DOMAIN 1479 1577 C2 DOMAIN 4.
 FT DOMAIN 1305 1310 POLY-TYS.
 FT DOMAIN 1314 1320 POLY-GLU.
 FT DOMAIN 1970 1978 POLY-LEU.
 FT VASPLIC 169 169 R -> SKGRETEKGRGDEHK (IN ISOFORM 2).
 FT VASPLIC 1244 1263 MISSING (IN ISOFORM 2).
 FT VASPLIC 1943 1997 SEIPLNPLKSAFYPLMWTYVLLKFLFLILLALPL
 FT VASPLIC 1943 1997 YSLPGYLAKRTIGA -> AFWPLNPLKSKYILCTGRKRL
 FT IRIYVLLALGGLMALFLPLSLPGYVKKLLGA (IN
 FT ISOFORM 2).
 FT SEQUENCE 1997 AA: 227031 MW: 799CA918F79DAD1 CRC64;
 Query Match 39.1%; Score 61; DB 1; Length 1997;
 Best Local Similarity 1.3%; Pred. No. 3.5e+02;
 Matches 18; Conservative 1; Mismatches 6; Indels 1316; Gaps 4;

QY 7 SMEXR----- 11
 DB 113 SMEVRYGATDGYVPMWDGDFLDESLQEEKDSQETDGLLPGSRPSTRUSGKSFRRAGR 172
 QY 12 ----- 11
 DB 173 SVFSAMKLGKTRSHKEPQRQDEPAVLEMDLDHLAIQLGDLDPDSVLSAVTALTSNV 232
 QY 12 ----- 11
 DB 233 SNKRKPDIMPSASRPMQYQVSTVTEARQLVGLMDPVCEVEGDKKTKTSKKESTN 292
 QY 12 ----- 11
 DB 293 CPYYNEVFDFHVPDVMFDKILIKTSVHSKNILRSGTLVGSFKMDVGYVSQPEHQR 352
 QY 12 ----- 11
 DB 353 RKWALISDPDDISAGLKGVCDAVVGKGNIKTPHKANETDEDDIEGNLLPEGVPE 412
 QY 12 ----- 11
 DB 413 ROMARFYKIRAGLEPRMNTSLMANKAFIGENTDVPYQVFPAGCKTSYQAKS 472
 QY 12 -----TPDI 16
 DB 473 YEPLMNGVFTDLPPPLCKRMKQVQRDSKVNDVAIGTFIDLRKISNDGKGFLPTLC 532
 QY 17 PAM----- 19
 DB 533 PAMVNNATGSTRNYTLDEHQLNEGIGVSPFARMLGLAVILDTSPBELTSSTEVQV 592
 QY 20 ----- 19
 DB 593 EQATPVSECTGAKKEEFLGALFLEASKIDRKNDGPITPEVTIGATGVEYDGMNPLAP 652
 QY 20 ----- 19
 DB 653 RPRREPDEEEVDLIQNSDDEDEAGDLASVSTPPMRPQIUDRNYFHLPLERKPCIT 712
 QY 20 ----- 19
 DB 713 IKSMPDQRRRLYNANIMDIADKLEGLNDVQEMIKTEKSPERRLGLVLELSCGHR 772
 QY 20 ----- 19
 DB 773 FLSLSDKQGRSSRTLRDRRLKSCMRLESMGOAKSLPAQYKRRTVNDKRLCQNFQ 832
 QY 20 ----- 19
 DB 833 KLFIADEPQHSIPDVFIWMSNNKRIAYKAVPSKDLFSIVEELGKCAKVTLEFL 892
 QY 20 ----- 19
 DB 893 PGKRGFSAGVQAKIELTYLTGLSKQRDFTGLPGCFEEYKAAQGLSLSPISLV 952
 QY 20 -----YXKR----- 23
 DB 953 YTKKQAFQLAHNTQARSLFPAADSSGLSDPFAVFTINOSCTEVANETLCPTQDMVVF 1012
 QY 24 ----- 23
 DB 1013 DNLELGEAHELDPPILVIEIYDQDSMKADPMGRTEAKPLVKKADEAYCPRPPOL 1072
 QY 24 -----GIRP----- 27
 DB 1073 EYQIYNGSATAGDLLAFAELLQIGPSGRADLPPIINGPVMQNGPIPVYVGIRPVLSTK 1132
 QY 28 ----- 27
 DB 1133 RVEVLFNGLDDLKRVNLAQVDRPRVDIECAGKGVQSSLIHNTKANPNTLVKRVFVDLP 1192

QY 29 ----- 28
 DB 635 VQHDKKEPVVYSPYSDHFLRMGVADNGEPIDRQIKYCPYKISGTWLENSCMTVE 694
 QY 29 ----- 28
 DB 695 VNETSPENTOLVGNTRYRIELKANAIGYSPASIMKTKGIDVIGVAERQVSSAAI 754
 QY 29 ----- 28
 DB 755 VGIAIGVLLLFVYDILGCTIVHGVNATMCRKAKRSPSEIDDEAKIGSGQVKEPPPS 814
 QY 29 ----- GRF 31
 DB 815 PLEPPVKGSGPMSTPLEDEKPLRPTGSIKONSTIEFGRR 858
 RESULT 11
 DVRA_METTB STANDARD: PRT: 962 AA.
 ID DVRA_METTB
 AC 026543;
 DF 30-MAY-2000 (Rel. 39, Created)
 DF 30-MAY-2000 (Rel. 39, Last sequence update)
 DF 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exonuclease ABC subunit A.
 GN DVRA OR MTR443.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 CC Methanobacteriaceae; Methanothermobacter.
 CX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubols J.,
 Airdedge T., Bashirzadeh R., Blakely D., Cook R., Gilbert R.,
 Harrison D., Hoang L., Keagle P., Lumm W., Pochter B., Qiu D.,
 Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 Daniels C.J., Mac J.-I., Rice P., Noelling J., Reeve J.N.;
 RA *Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7153(1997).
 CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
 CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). DVRA IS AN ATPASE
 CC AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
 CC STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS: DVRA, DVRA AND DVRC.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. DVRA SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AE000828; AAB84949.1.
 DR InterPro: IPR004349; ABC_transportr.
 DR InterPro: IPR004602; DVRA.
 DR Pfam: PF000005; ABC_tran. 2.
 DR ProDom: PD000606; ABC_transportr. 1.
 DR TIGRfam: TIGR00630; dvra. 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER. 2.
 KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
 KW DNA-binding; Zinc-finger; Complete proteome.
 FT NP_BIND 38 45 ATP (POTENTIAL).
 FT NP_BIND 649 656 ATP (POTENTIAL).
 FT ZN_FING 748 774 C4-TYPE.

SQ SEQUENCE 962 AA; 108395 MW; 2C0EF7FC41CCD060 CRC64;
 Query Match 38.5%; Score 60; DB 1; Length 962;
 Best Local Similarity 3.18; Pred. No. 1.2e+02;
 Matches 17; Conservative 2; Mismatches 10; Indels 51; Gaps 3;
 QY 2 2RXHSHNE----- 9
 DB 201 RNRKHSIDVVVDRLVYVRDTEFRKRLADSVETALQEGTVRVPSTMTGGEETISEHRA 260
 QY 10 ----- 9
 DB 261 CPETAGINFEIISPRMFSPNSPACPCNGCGSKLEIDPDLYVPYPSINEGALVPS 320
 QY 10 ----- 9
 DB 321 KSKRDYTYQMLRAVAEHGFSLDTPFRDDEHRRALITGDEKIGVFQKNTYR 380
 QY 10 ----- 9
 DB 381 NRREGVIPMERIYMETKSNYRTYIGREMSHACPVCGSRIRPESLVTNGSIHD 440
 QY 10 ----- 9
 DB 441 VVEMSTREAHFPDLSKLTREERYIAREVLKETRELRFLIDVGLDYLTLSRSSGTLSCG 500
 QY 10 ----- 9
 DB 501 EAQRIRATIGSGIYGLVILDEPSIGLHQRRNRRLIETLRKRLDGLMTLYVEHDEET 560
 QY 10 ----- 9
 DB 561 ILSADHVVDIGPGAGHGCVVAGTPEELMEDPDSLTGALYSRETIPLPEVRRRPSGR 620
 QY 10 ----- 9
 DB 621 YLYRGAENNLKELIDVRLTGLFCTYGVSGGSKTYNDILRGYERLNKHNACR 680
 QY 10 ----- XRTPDINPMYX----- XRGIRPVGR 31
 DB 681 HDIEGLQHDIKVMIDPSIGRTPSNPATYGVTFHIELPQAPPEARRRGYR-GRF 739
 RESULT 12
 HRX_MOUSE STANDARD: PRT: 3866 AA.
 ID HRX_MOUSE
 AC P55200;
 DF 01-OCT-1996 (Rel. 34, Created)
 DF 01-OCT-1996 (Rel. 34, Last sequence update)
 DF 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein HRX (ALX-1) (Fragment).
 GN MLL OR HRX OR ALL1.
 OS Mus musculus (mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; and C57BL/6 x CBA; TISSUE=Spleen, and Lung;
 RX MEDLINE=93317679; PubMed=8327517;
 RA Ma O., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
 Chanaani E., Croce C.M., Stracusa L.D., Buchberg A.M.;
 RA "Analysis of the murine All-1 gene reveals conserved domains with
 RT human ALL-1 and identifies a motif shared with DNA
 RT methyltransferases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
 CC EMBRYONIC DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.


```

CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: L17069: AAA62593.1; -.
DR MGD: KGI:96995; M1.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003889; FYRICH.
DR InterPro: IPR003888; FYRICH_N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR002857; Zn1_CXXC.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50280; SET; 1.
DR PROSITE: PS01359; ZF_PHD_1; 3.
DR PROSITE: PS50016; ZF_PHD_2; 3.
DR DNA-binding: Nuclear protein; Zinc-finger; Metal-binding;
DR Transcription regulation; Alternative splicing; Polymorphism.
FT NON_TER 1
FT DNA_BIND 67 78 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 115 125 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 199 207 A.T HOOK (BY SIMILARITY).
FT ZN_FING 1044 1091 CXXC-TYPE.
FT ZN_FING 1330 1381 PHD-TYPE 1.
FT ZN_FING 1383 1432 PHD-TYPE 2.
FT ZN_FING 1465 1529 PHD-TYPE 3.
FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).
FT DOMAIN 3737 3846 SET.
FT DOMAIN 35 41 POLY-GLY.
FT DOMAIN 459 469 PRO-RICH.
FT DOMAIN 1231 1238 POLY-PRO.
FT DOMAIN 3533 3536 POLY-GLU.
FT VARSPLIC 3693 3697 POLY-GLU.
FT VARSPLIC 1503 1505 MISSING (IN ISOFORM 2).
FT VARIANT 1497 1497 K -> T.
SQ SEQUENCE 3666 AA; 420976 MW; ADRC55E14E806F1D CR664;

Query Match 36.5%; Score 60; DB I; Length 3666;
Beat local similarity 0.5%; Pred. No. 1.4e+03;
Matches 15; Conservative 6; Mismatches 9; Indels 2968; Gaps 4;
QY 2 RXHXSMEXRT----- 12
DB 647 RSPSHSMRTSGRLSTSELSPITPPSVSSLSIFVSPLASALNPITFPSSHILTQSGD 706
QY I3 ----- 12
DB 707 STEKNRANKQTSAPAEPPSSNPALFPWFTPGSQTEKGRKKDPAPELSKDRADKSYE 766
QY I3 ----- 12
DB 767 KDKSREREREKEHKKRKRKGGSDIQSSALYPVGRVSKKRVAGEDVTSSAKK 826
QY I3 ----- 12

```

```

DB 827 ATGRKKSSLDGADVAPVTLLGTTAVAKATILKGRGNLEKNLIDGPAAPSLERERP 886
QY I3 ----- 12
DB 887 CLSAPSSSYVHSTSSIGSMIAQADKLPMTDKRYASLLKAKAKQICKTEKSLKQTDQ 946
QY I3 ----- 12
DB 947 KAQGESDSESVRGPRIKHYBRAVALGRKRAVEPDDMTLSALPWEREKILSMG 1006
QY I3 ----- 12
DB 1007 NDKRSVAGSEDAEPLAPPIKPIKVTNRNKAPOEPPVYKGRSRRCQCPGQVPEDCGI 1066
QY I3 ----- 12
DB 1067 CTNCLDKPRKGGRIKQCKMKRCQNLQWMPKASIQKQTKAVKKKSKTEKESK 1126
QY I3 ----- 12
DB 1127 ESTSVKSPLEPAQAAPPREPAFKSSSEPPRKVEEKESEGGAPAPAPAEPRQYS 1186
QY I3 ----- 12
DB 1187 APASRKKSKQVOPAAVVPPOPTAPQKKAAPSEPKKQPPPEPGPQSKQKV 1246
QY I3 ----- 12
DB 1247 APLSIVYKQPKDKKPPVYSQENAGTILNPLNSIGSSKQIPADGVHRIKRVDFKE 1306
QY I3 ----- 12
DB 1307 DCAENYEMWGLILTSVPIIPRVCFCLSSBHYEYVQYCEPPEFKFLEHERPL 1366
QY I3 ----- 12
DB 1367 EDLENNCCRCKKCHVCGROHQTQQLLECNCRNSYHECIGPNYPTFKKKKWIIC 1426
QY I3 ----- 12
DB 1427 TKCYRCSGSGTTPGCKMDQMSHDSLCHDCAKIFAKGFCPLCDKCYDDDYESKMQ 1486
QY I3 ----- 12
DB 1487 CGKCDRWVHSKESLSTGEDEMYEILSNLPESVAYTCVCTERHPPKRLALEKEIQASL 1546
QY I3 ----- 17
DB 1547 KQVITALINSHETSHLLRYQAAKPPDLNPEDESIPSRSSPGGPPVLTVEVKODEQ 1606
QY I8 ----- 17
DB 1607 PLDEGKAKKNDQGSYVLEFSDDIYKTIQAAINSIDGQPELIKANSVKSFFIQMER 1666
QY I8 -ANYXK------ 23
DB 1667 VFPEFSVKKSHFEPNKNVSNNGSLPNAVLPPSLDHYAQMQRRESSNTEQPLMKTI 1726
QY I24 ----- 23
DB 1727 PAPKPKGPEEDSPTLHPPTPTLSTDRSREDSPELNPBGIDNRQALCLMYGDDSA 1786
QY I24 ----- 23
DB 1787 NDAGRLIYIGQNEHTHVNCALMSAEVEEDDGSILKNVHAIVINGKQLRCERQPGATVG 1846
QY I24 ----- 23
DB 1847 CCLISTSTNYHFMCSRAKNCVFLDDKKRYVCQRHDLIKGEVPENGEYVRVYDFEGL 1906
QY I24 ----- 23

```

```

Db 1907 SLRRKFLNGLEPENIHMGISMTIDCIGILNDLSDCEDKLEPIGYQCSRYVSTTDARKR 1966
QY 24 ----- 23
Db 1967 CVYCKIMECRPPVEPDINSTVEHDDNRTIANSPSSIDASCDSOATAILSPSPDR 2066
QY 24 ----- 23
Db 2027 PHSOTSGCYHYHISKVPRINTPSYPTQSPGCRPLPSAGSPPTHEIVTVGDPLISS 2086
QY 24 GIRPVG----- 29
Db 2087 GLRISGRNRHSTSLSLRSLKRLINSPYRTGSANRSGSVSPSLGTATPDENAKASADR 2146
QY 30 ----- 29
Db 2147 GGLSSANLGHSAFPSSSGRTVGSGTSHLDGSPSEVAKSALDLVPKSLVKRKN 2206
QY 30 ----- 29
Db 2207 RTSSSKTDSGSAHSTAYPGIKPLTPOVNRNATPGELINISKIGFAEPSTVPSSKDTVSTP 2266
QY 30 ----- 29
Db 2287 QLHLRGQRSDQKHNDPSQSVKPSPNEDGEIKTILKLPGMHRPSILHEHIGSSRDRRK 2326
QY 30 ----- 29
Db 2327 GKRSKETCKEKHSKSYLERQVYTTGEGNLKPEFADEVLTTPFLGQRPNNTYSRKIG 2386
QY 30 ----- 29
Db 2387 DKVLPJLPGVPKQSTQVEGSKKELQAPRKCSVKYPLKMEGENSKNTQKESGSPSAHI 2446
QY 30 ----- 29
Db 2447 ESVCPAEVSASRBPAGAPCVQPPNNLTLSQDPPQNNNTQNLPEQDRNLMLPDGKPEQEDG 2506
QY 30 ----- 29
Db 2507 SFKRPRRSARARSNNMFFGLTPLYGVMSYGEDIPFYSNSTGKRGKRSAGQVGDGAD 2566
QY 30 ----- 29
Db 2567 LSTDEDDLYYNTFRVYISSGGERLASHNLFREEQCDLPKISOLDVDDGTESDTSV 2626
QY 30 ----- 29
Db 2627 TATSRKSSQIKRKMKENGENTMLKIDREDAKEEHYKSAVGHAKNEPKILDNCHSVSRVK 2686
QY 30 ----- 29
Db 2687 AAGQDLEAQLSSLESSRVRHISTPSDKNLLDTYNAELKSDSDNNNSDDCGNLLPSDIM 2746
QY 30 ----- 29
Db 2747 DEVLKNTPSMQLGESPESSSELLTGEGLGLDNNEKDIGLEVESQQLPATEPYDSS 2806
QY 30 ----- 29
Db 2807 VSSSI SAEQFELPLEPSDL SVLTTSPTVPSQNPRLAVISDGRKRVITTEKSVASS 2866
QY 30 ----- 29
Db 2867 EGDPAALSPGVADPEGHMTPDHF LQGHMADHISPPCGSVEQGHNSQDLTRNSGTPG 2926
QY 30 ----- 29
Db 2927 LQVPVSPVQONOKYVPSSTDSGSPQISNAVAQTTPHLKPATEKLIYVQNMQPLVY 2986
QY 30 ----- 29
Db 2987 LQTLPGVYVQKIQLTSPVSSTPSVMEHTVSVLGPMSGLITLTGLNPSLPPSPSLPPAS 3046

```

```

QY 30 ----- 29
Db 3047 KGLLSVPHHQLHLSFPAAAGSSPPRVISSPPSGLLIGVPPDPQLGSEANQRIDLTIT 3106
QY 30 ----- 29
Db 3107 VAPSSGLKRPRIURLHTRKNNKILAPSSAPSNIAVSQVSNMFLINFPQSLSNHPSLID 3166
QY 30 ----- 29
Db 3167 LGSLSNPSHRTVFNILIKRSGKIMFEPQAPLLPPOSVGSTAAFAAGSSTISODTSHLTSG 3226
QY 30 ----- 29
Db 3227 PVSAALSGSSVLNVSMQTAAFTSSTVPGHYTLANQRLGTPDIGSHLLIKASHOS 3286
QY 30 ----- 29
Db 3287 LGIQDPVALPPSSGNFPOLGTSQTPSAAMTAASGICVLPSSTQTAGMTAASPGEAEERH 3346
QY 30 ----- 29
Db 3347 YKLQGNQLLAGKTGTLTISQRDRDPDAPGTPSTFTQTAEPNGVSLFQNTLPKAPPA 3406
QY 30 ----- 29
Db 3407 SSASPGSSPSSGQSGSSSVPGPTKPKAKRRLQLPDKSVKHKHVSHLRTSSSAHTPH 3466
QY 30 ----- 29
Db 3467 RQDPAPOPSTVTRPRANEQDDAGVEOPSOKEGQAPPAVALPEVQATQNPANBQEN 3526
QY 30 ----- 29
Db 3527 AEPKAMEEESGSSPLMLMLQOEKREKSTERRKPKGLVEIISDDGFQICASIEDA 3586
QY 30 ----- 31
Db 3587 WKSLTDKVEARSNARLKLQLSFAGVNGLRMLGLMDAVVFLIEQLAGAKHCRNYKRF 3644

```

RESULT 13
 HRX_HUMAN
 ID HRX_HUMAN STANDARD; PRT: 3969 AA.
 AC Q03164; Q14845; Q16364; Q13743; Q13744; Q9UW43;
 DT 01-OCT-1993 (Rel. 27, Created)
 DF 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein HRX (ALU-1) (trithorax-like protein).
 GN MIL OR HRX OR ALU1 OR TRX1 OR RTRX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93046667; PubMed=1423624;
 RA Tkachuk D.C., Kohler S., Cleary M.L.;
 RT "Involvement of a homolog of Drosophila trithorax by 11q23
 RT chromosomal translocations in acute leukemias.";
 RL Cell 71:691-700(1992).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96290553; PubMed=8703835;
 RA Nilson I., Loechner K., Stiegler G., Grell J., Beck J.D., Fey G.H.,
 RA Marschalek R.;
 RT "Exon/intron structure of the human ALU-1 (MLL) gene involved in
 RT translocations to chromosomal region 11q23 and acute leukaemias.";
 RL Br. J. Haematol. 93:966-972(1996).
 RN 13
 RP SEQUENCE OF 1-1909 FROM N.A.
 RX MEDLINE=93390935; PubMed=8378076;

RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
 RA Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda K.,
 RT "Two distinct portions of LTRg1/ENL at 19p13 are involved in t(11;19)
 RT leukemia.";
 RT Oncogene 8:2617-2625(1993).
 [4]
 RP SEQUENCE OF 1317-2328 FROM N.A.
 RP TISSUE-BRAIN.
 RX MEDLINE-93265134; PubMed-1303259;
 RA Djabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.,
 RT "A trithorax-like gene is interrupted by chromosome 11q23
 RT translocations in acute leukaemias.";
 RT Nat. Genet. 2:113-118(1992).
 [5]
 RP SEQUENCE OF 1251-1538 FROM N.A.
 RP MEDLINE-94215165; PubMed-8162575;
 RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaan O.,
 RA Saito H., Croce C.M., Canaan E.;
 RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
 RT involved in acute leukemia.";
 RT Cancer Res. 54:2326-2330(1994).
 [6]
 RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
 RP MEDLINE-95322025; PubMed-7598802;
 RA Mbangkollo D., Burnett R., McCabe N., Thirman M., Gili H., Yu H.,
 RA Rowley J.D., Diaz M.O.;
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
 RT trz zinc-finger domain, and alternative splicing.";
 RT DNA Cell Biol. 14:475-483(1995).
 [7]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RP MEDLINE-95315013; PubMed-7794749;
 RA Marschalek R., Grell J., Lochner K., Nilson I., Steigler G.,
 RA Zechner I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
 RT translocation t(4;11).";
 RT Br. J. Haematol. 90:308-320(1995).
 [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RP MEDLINE-94020842; PubMed-841518;
 RA Forster A., Rabbits T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 RT application to isolation of MLL fusion cDNAs from acute leukaemia
 RT translocations.";
 RT Oncogene 8:3157-3160(1993).
 [9]
 RP CHROMOSOMAL TRANSLOCATION WITH GAST.
 RP MEDLINE-20183971; PubMed-10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Wilson R.B., Jones D.H., Adjei K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.H., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAST translocation early during
 RT chemotherapy with DNA topoisomerase II inhibitors.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 [10]
 RP CHROMOSOMAL TRANSLOCATION WITH AFBP21.
 RP MEDLINE-20115194; PubMed-10648423;
 RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
 RT "Novel SH3 protein encoded by the AFBP21 gene is fused to the mixed
 RT lineage leukemia protein in a therapy-related leukemia with
 RT t(3;11)(p21;q23).";
 RT Blood 95:1066-1068(2000).
 [11]
 RP FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 RP SUBCELLULAR LOCATION: Nucleus.
 [12]
 RP TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 [13]
 RP DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL;
 T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
 THAT INVOLVES MLL AND MLLT3/AF6; T(6;11)(Q27;Q23) THAT INVOLVES
 MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
 MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;
 T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;

CC T(11;19)(Q21;Q23) THAT INVOLVES MLL AND AFBP21; T(11;19)(Q23;P13.3)
 CC that involves MLL and ENL; t(11;19)(q23;p23) that involves MLL
 CC and GAST; and t(3;11)(p21;q23) that involves MLL and AFBP21.
 CC - SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC - SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC - SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC - SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC - SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC - DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/genes/MLL.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL: L04284; AAA58669.1; -
 DR EMBL: 269744; CAA93625.1; -
 DR EMBL: 269745; CAA93625.1; JOINED.
 DR EMBL: 269746; CAA93625.1; JOINED.
 DR EMBL: 269747; CAA93625.1; JOINED.
 DR EMBL: 269748; CAA93625.1; JOINED.
 DR EMBL: 269749; CAA93625.1; JOINED.
 DR EMBL: 269750; CAA93625.1; JOINED.
 DR EMBL: 269751; CAA93625.1; JOINED.
 DR EMBL: 269752; CAA93625.1; JOINED.
 DR EMBL: 269753; CAA93625.1; JOINED.
 DR EMBL: 269754; CAA93625.1; JOINED.
 DR EMBL: 269755; CAA93625.1; JOINED.
 DR EMBL: 269756; CAA93625.1; JOINED.
 DR EMBL: 269757; CAA93625.1; JOINED.
 DR EMBL: 269758; CAA93625.1; JOINED.
 DR EMBL: 269759; CAA93625.1; JOINED.
 DR EMBL: 269760; CAA93625.1; JOINED.
 DR EMBL: 269761; CAA93625.1; JOINED.
 DR EMBL: 269762; CAA93625.1; JOINED.
 DR EMBL: 269763; CAA93625.1; JOINED.
 DR EMBL: 269764; CAA93625.1; JOINED.
 DR EMBL: 269765; CAA93625.1; JOINED.
 DR EMBL: 269766; CAA93625.1; JOINED.
 DR EMBL: 269767; CAA93625.1; JOINED.
 DR EMBL: 269768; CAA93625.1; JOINED.
 DR EMBL: 269769; CAA93625.1; JOINED.
 DR EMBL: 269770; CAA93625.1; JOINED.
 DR EMBL: 269771; CAA93625.1; JOINED.
 DR EMBL: 269772; CAA93625.1; JOINED.
 DR EMBL: 269773; CAA93625.1; JOINED.
 DR EMBL: 269774; CAA93625.1; JOINED.
 DR EMBL: 269775; CAA93625.1; JOINED.
 DR EMBL: 269776; CAA93625.1; JOINED.
 DR EMBL: 269777; CAA93625.1; JOINED.
 DR EMBL: 269778; CAA93625.1; JOINED.
 DR EMBL: 269779; CAA93625.1; JOINED.
 DR EMBL: 269780; CAA93625.1; JOINED.
 DR EMBL: D14540; BAA03407.1; -
 DR EMBL: L01986; AAA92511.1; -
 DR EMBL: D04737; AAA18644.1; -
 DR EMBL: S78570; AAA34770.1; -
 DR EMBL: X83604; CAA58584.1; -
 DR EMBL: S66432; AAA28345.1; -
 DR EMBL: AF23198; AAG26332.2; ALT_TERM.
 DR TRANSFAC: T02337; -
 DR Genew: HGNC:7132; MLL.
 DR MIM: 159555; -
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR003889; Fitch_C.
 DR InterPro: IPR003888; Fitch_N.
 DR InterPro: IPR003616; PostSet.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR002857; Znf_CXXC.
 DR InterPro: IPR001965; Znf_PHD.


```

QY 30 ----- 29
Db 2489 DKGLSMGVKAPPMQVGSAAKELQAPRKCTVAVTLTPLMEENSQSKALKESSPASPL 2348
QY 30 ----- 29
Db 2349 QLESTPTEPISASENFGDGVAPSPNNTSCQDOSQNNQNLVQDRNLMLDPGRKPE 2608
QY 30 ----- 29
Db 2609 DGEFKRRYPRRSARASNMFGTLPLYVRSYGEEDIPYSSYTKKKGRSAEGYDGA 2668
QY 30 ----- 29
Db 2669 DDLSTDEDLIYNTFRITYISSGGERLASHNLFREEQDLPKISQLDGVDDGESDT 2728
QY 30 ----- 29
Db 2729 SVATTKRSQIPKRNKGKNGENTLAKIDPEADAGEKEHYTKSSVGHKNEPKMNCISVR 2788
QY 30 ----- 29
Db 2789 VKTGGDSLEAQLSLESSRRVHTSPSPDKNLDITYNTLKLSDNNNSDDCGNLPD 2848
QY 30 ----- 29
Db 2849 IKDFVLKNTPSMAAGSPSSSELLNLGEGLGDSNREKMGLEFVSQQLPTEPVD 2908
QY 30 ----- 29
Db 2909 SSVSSISASEQFELPELPDLSVLTRSPYPSQNSRLAVISDGEKRYITTKSVA 2968
QY 30 ----- 29
Db 2969 SSESDFALLSPGVDPPPEGHMTPDHFIQGMADHISPPCGSVEGHGNNQDLIRNSST 3028
QY 30 ----- 29
Db 3029 PGLQVPSPYPIQNKQYVNSTDSPPSOISNAQVOTTPPNLKATEKLIYVQNMQL 3088
QY 30 ----- 29
Db 3089 YVLQTLNPGVTKIQLTSSVSPSVAVETWTVLGPWGGLTITGLNPLSLTQSLEPS 3148
QY 30 ----- 29
Db 3149 ASKGLPMKSHHQLNSFPAAQTSSPSPNISNPSGLLIGVPPDPQLVSESSORTDLS 3208
QY 30 ----- 29
Db 3209 TVVATPSSGLKRRPISRLQTRKKNKLAPSSPSHAPSDVVSNNMLINTPSQLPNNPSL 3268
QY 30 ----- 29
Db 3269 LDGLSNTSSNRVVPNIIRKSSIMWFEPAPLIPQSVGTATAGTSPISQDTHLS 3328
QY 30 ----- 29
Db 3329 GSVGLASSSVLVNWSMOTTTPTSSASVPHVTLINPELLGTPDIGSINLLIRASQ 3388
QY 30 ----- 29
Db 3389 SLGIQDPVVALPSSSGMFPQLGTSQPTPAITAASSICVLPSTQTTGITAASPSGEADE 3448
QY 30 ----- 29
Db 3449 HYOLAHVNLILASKTGRSSORDIDASGPQVSNFTQVDAPNSMGLQNKALSAVQAS 3508
QY 30 ----- RF 31
Db 3509 PTPGSGSPSSSGQSRASPSVPGPTKPKTKRF 3543

```

```

RESULT 14
ID PGEM_HUMAN STANDARD: PRI: 4393 AA.
AC P98160; Q16287;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Basament membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92112994; PubMed-1730768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor."
RT J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin, and Colon;
RX MEDLINE-92235084; PubMed-1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor."
RT J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 1018-1472 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE-91365376; PubMed-1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1."
RT Genomics 10:673-680(1991).
RN [4]
RP SEQUENCE OF 892-1398 FROM N.A.
RC TISSUE=Fibroblasts;
RX MEDLINE-92120660; PubMed-1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1-1p35 and identification of
RT a BamHI restriction fragment length polymorphism."
RT Genomics 11:389-396(1991).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE-94052171; PubMed-8234307;
RA Cohen I.R., Graessell S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter."
RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [6]
RP FUNCTION: This protein is an integral component of basement
RN membranes. It is responsible for the fixed negative electrostatic
RN charge and is involved in the charge-selective ultrafiltration
RN properties. It serves as an attachment substrate for cells.
RN [7]
RP SUBUNIT: Purified perlecan has a strong tendency to aggregate in
RN dimers or stellate structures. It interacts with other basement
RN membrane components such as laminin, prolargin and collagen type
RN IV.
RN [8]
RP SUBCELLULAR LOCATION: Extracellular.
RN [9]
RP TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
RN [10]
RP PTM: CONTAINS THREE NEPARAN SULFATE CHAINS AS WELL AS N-LINKED
RN AND O-LINKED OLIGOSACCHARIDES.
RN [11]
RP SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.

```

CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X62515; CAA44373.1; -;
 DR EMBL: M65289; AAS2700.1; -;
 DR EMBL: M64283; AAS2699.1; -;
 DR EMBL: S76436; AAB21121.2; -;
 DR HSP: L22078; -; NOT_ANNOTATED_CDS.
 DR HSP: P00740; IFDM.
 DR Stena-2DPAGE; P98160; -;
 DR Genew; HGNC:3273; HSPG2.
 DR MIM; 142461; -;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF 2.
 DR InterPro: IPR001438; EGF II.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR000034; laminin_B.
 DR InterPro: IPR002049; laminin_EGF.
 DR InterPro: IPR001791; laminin_G.
 DR InterPro: IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00047; Ig; 22.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 7.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; Idl_recept_a; 4.
 DR Pfam; PF03390; SEA; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR ProDom; PD003031; laminin_B; 3.
 DR SMART; SM00180; EGF_Iam; 6.
 DR SMART; SM00408; EGF_Like; 8.
 DR SMART; SM00192; IGc2; 22.
 DR SMART; SM00281; LDla; 4.
 DR SMART; SM00281; Iamg; 3.
 DR SMART; SM00282; Iamg; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS50025; Iam_G_DOMAIN; 3.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS50024; SEA; 1.
 DR Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain;
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 4393 BASEMENT MEMBRANE-SPECIFIC REPARAN
 FT DOMAIN 80 194 SULFATE PROTEOGLYCAN CORE PROTEIN.
 FT DOMAIN 197 236 SEA.
 FT DOMAIN 283 321 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 323 361 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 366 405 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 405 506 IG-LIKE C2-TYPE DOMAIN I.
 FT DOMAIN 523 532 LAMININ EGF-LIKE 1 (N-TERMINAL).
 FT DOMAIN 533 732 LAMININ EGF-LIKE 1 (DOMAIN III A).
 FT DOMAIN 733 765 LAMININ EGF-LIKE 1 (C-TERMINAL).
 FT DOMAIN 766 815 LAMININ EGF-LIKE 2.

FT DOMAIN 816 873 LAMININ EGF-LIKE 3.
 FT DOMAIN 881 925 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 926 935 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 936 1127 LAMININ DOMAIN IV 2 (DOMAIN III B).
 FT DOMAIN 1128 1160 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 1161 1210 LAMININ EGF-LIKE 6.
 FT DOMAIN 1211 1267 LAMININ EGF-LIKE 7.
 FT DOMAIN 1277 1336 LAMININ EGF-LIKE 8.
 FT DOMAIN 1337 1336 LAMININ EGF-LIKE 9 (N-TERMINAL).
 FT DOMAIN 1337 1531 LAMININ EGF-LIKE 9 (C-TERMINAL).
 FT DOMAIN 1532 1564 LAMININ EGF-LIKE 9 (C-TERMINAL).
 FT DOMAIN 1565 1614 LAMININ EGF-LIKE 11.
 FT DOMAIN 1615 1672 LAMININ EGF-LIKE 11.
 FT DOMAIN 1679 1773 LAMININ EGF-LIKE 11.
 FT DOMAIN 1774 1867 LAMININ EGF-LIKE 11.
 FT DOMAIN 1868 1957 LAMININ EGF-LIKE 11.
 FT DOMAIN 1958 2053 LAMININ EGF-LIKE 11.
 FT DOMAIN 2054 2153 LAMININ EGF-LIKE 11.
 FT DOMAIN 2154 2246 LAMININ EGF-LIKE 11.
 FT DOMAIN 2247 2342 LAMININ EGF-LIKE 11.
 FT DOMAIN 2343 2438 LAMININ EGF-LIKE 11.
 FT DOMAIN 2439 2535 LAMININ EGF-LIKE 11.
 FT DOMAIN 2536 2631 LAMININ EGF-LIKE 11.
 FT DOMAIN 2632 2728 LAMININ EGF-LIKE 11.
 FT DOMAIN 2729 2828 LAMININ EGF-LIKE 11.
 FT DOMAIN 2829 2926 LAMININ EGF-LIKE 11.
 FT DOMAIN 2927 3023 LAMININ EGF-LIKE 11.
 FT DOMAIN 3024 3114 LAMININ EGF-LIKE 11.
 FT DOMAIN 3115 3213 LAMININ EGF-LIKE 11.
 FT DOMAIN 3214 3300 LAMININ EGF-LIKE 11.
 FT DOMAIN 3301 3401 LAMININ EGF-LIKE 11.
 FT DOMAIN 3402 3490 LAMININ EGF-LIKE 11.
 FT DOMAIN 3491 3576 LAMININ EGF-LIKE 11.
 FT DOMAIN 3577 3654 LAMININ EGF-LIKE 11.
 FT DOMAIN 3655 3845 LAMININ EGF-LIKE 11.
 FT DOMAIN 3846 3883 LAMININ EGF-LIKE 11.
 FT DOMAIN 3884 3924 LAMININ EGF-LIKE 11.
 FT DOMAIN 3925 4105 LAMININ EGF-LIKE 11.
 FT DOMAIN 4106 4143 LAMININ EGF-LIKE 11.
 FT DOMAIN 4145 4178 LAMININ EGF-LIKE 11.
 FT DOMAIN 4203 4391 LAMININ EGF-LIKE 11.
 FT SITE 65 67 EGF-LIKE 1.
 FT SITE 71 73 EGF-LIKE 2.
 FT SITE 76 78 EGF-LIKE 3.
 FT SITE 4151 4153 EGF-LIKE 4.
 FT SITE 4301 4303 EGF-LIKE 5.
 FT DISULFID 199 212 LAMININ G-LIKE 3.
 FT DISULFID 206 225 LAMININ G-LIKE 3.
 FT DISULFID 219 234 LAMININ G-LIKE 3.
 FT DISULFID 265 297 LAMININ G-LIKE 3.
 FT DISULFID 292 310 LAMININ G-LIKE 3.
 FT DISULFID 304 319 LAMININ G-LIKE 3.
 FT DISULFID 325 337 LAMININ G-LIKE 3.
 FT DISULFID 332 350 LAMININ G-LIKE 3.

Query Match 38.5%; Score 60; DB 1; Length 4393;
 Best Local Similarity 0.5%; Pred. No. 1; 8e+03;
 Matches 19; Conservative 3; Mismatches 9; Indels 3672; Gaps 5;

OY 1 SRX-----
 DB 648 SRCHTPTOPCALNQRQVQSEEHVHESGRPVQARLQVLASLEAVLIQVYNTKMASV 707
 OY 5 -----XHSME-----
 DB 708 GLSDIAMDTVTWATSHGRAHSEVERCPDYGSGLSCESDAHFTVPDGPYLGTCSCGS 767
 OY 10 -----
 DB 768 CNGMASSCPVYGHCLNQHNTBGPQCKKACGFQDAKAKATATSCPCPYIDASRRF 827

QY 10 ----- 9
Db 828 SPYCLFDIDGQAICDCAKAPGYGRRCSCAPGEYGNPIQPGKCAPVNOEIVRCDERGS 887
QY 10 ----- 9
Db 888 GTSGEACRCKNNVGRICNECADRSFHSTNNPDCLKCFCKGVSRRCTSSWSRAQHLG 947
QY 10 ----- 9
Db 948 ASERGHFSLTNASTHTTNEGIFSPTEBELGFSFHHLLSGPTFMSLPRLGDKVTISY 1007
QY 10 ----- 9
Db 1008 GELRFTVQRSQPGSTPLHGQPLVYLQGNILHEHVAQEPSPQSTFIYPRQAOAQ 1067
QY 10 ----- 9
Db 1068 RPDGPATREHLLMAGIDTLIRASTAQAPAESRVSGISMDVAVPETGDPALLEVRQ 1127
QY 10 ----- 9
Db 1128 CSCPRGPGSCODDGTGTRTPSGLYLGTCSRCSCHGSEACEPRTGACQCGQHTTBSP 1187
QY 10 ----- 9
Db 1188 RCEOCPGYGAQNGSTPODCQLPCYGDPAAGQAHTCFLDTDGHPICDASPGHSGRH 1247
QY 10 ----- 9
Db 1248 CERCAFGYGNBSQGPCCORDSOVGP IGCNCDPGSVSSQCDMAQCCCKAQVEGLTCS 1307
QY 10 ----- 9
Db 1308 HCAPHFFLSASNPDCLEPCFKGITQCGASSAYTRHLSTHFPBGDFGFLVNPQRNS 1367
QY 10 ----- 9
Db 1368 RLNGEYVEVPRGAQLSPGNFAQLGHEFWOLPETYQDKVAAVAGKRLRYTLSTYAGP 1427
QY 10 ----- 9
Db 1428 QGSPLEDPDVOYITGNINMLVASOPALQGPERSYEIMFREEFWRPDGPATREHLLMAL 1487
QY 10 ----- 9
Db 1488 ADDELIRATFSSVPLVASISAVLEVAQGPSPNRPALVEYECNCPGYIGLSQDCA 1547
QY 10 ----- 9
Db 1548 PGYTRGSGLYLGHCELCRCNGHSDLCHEPTGACSOQHNNAEFCELCAPGYGDATAG 1607
QY 10 ----- 9
Db 1608 TPBDQPCACPLTNPENMSKTCESLAGAGYCTACEPGITGYCQCGRYVGNPSTVG 1667
QY 10 ----- 9
Db 1668 GQGLPETNOAPLVVEYHPARSIVPGGSHSLRCQVSGRPHFYFWRREDGRPVSTQOR 1727
QY 10 ----- 9
Db 1728 HOGSELHPFSVQPSDAGVYICTCRNLHRSNTSPABELLVT EABSKPIYVEEQRQSVPAR 1787
QY 10 ----- 9
Db 1788 GADVTFLCTAKKSPAYTLVWYRLANGKLPTRAMDNGILLITRANQOLSAGTYCTGSNA 1847
QY 10 ----- 9
Db 1848 FAMDGATLHWQAGSLASAVVSIHPQULTVQGLAEFRCSATSPITLEWTTGGPG 1907
QY 10 ----- 9

Db 1908 QLPKAIHGGILRLPAVEPTDQAQYLCAHSSAGQAVARVLHVGGGPRVQVSPERT 1967
QY 10 ----- 9
Db 1968 QVHAGRTVRLKCRAGVPSATITTRKRGSLPQOARSEPTDIALTLIPAITADAGYLC 2027
QY 10 ----- 9
Db 2028 VATSPAGTAQARIQVYVLSASDASQPPYKIESSPSVTBQOTDLNLCVAGSAHAQVTV 2087
QY 10 ----- 18
Db 2088 BRGSLPHHTQVNGSRULPQVSPADSGEYVCYVNGSGPREASITVSLGTHGSPSYT 2147
QY 19 ----- 18
Db 2148 FVPGSTRIRIEPSSSHVAEGQTLNLCVYHQAHAQVTHKRGSLPARHQTHGSLRL 2207
QY 19 ----- 18
Db 2208 HQVTPADSGEYVCHVYNGSGPLEASVLYTIEASVLPPIPEVRIESSSTVABQTLDS 2267
QY 19 ----- 18
Db 2268 CVVAGQAHAQVTVKRGSLPARHQVGRSLYIFQASPADAGQVCHASNGMEASITVTV 2327
QY 19 ----- 18
Db 2328 TGTGANTATPAGSTOPTRIEPPSSQVAEGQTLNLCVYHQAHAQVTHKRGSLPYRH 2387
QY 19 ----- 18
Db 2388 QTHGSLRLYQASPADSGEYVCYVNGSGPLEASVLYTIEPAGSPALGVPYVRIESS 2447
QY 19 ----- 18
Db 2448 SQVAGQTLNLCVAGQAHAQVTHKRGSLPARHQVGRSLRLLOYTPADSGEYVCYV 2507
QY 19 ----- 18
Db 2508 VGSSTQASVLYTIOQLSGSHSGQVAPVRIESSSASLANGTDLNLCVSAQAPHTI 2567
QY 19 ----- 18
Db 2568 TWYKRGSLPSRHQLYGSLRLPQVTPADSGEYVCHVNGASRSTSLIYTIQSGSSHV 2627
QY 19 ----- 18
Db 2628 PRVSPRIEISSPTVVBGQTLNLCVYARQPAITWYKRGSLPSRHQTHGSLRLHQ 2687
QY 19 ----- 18
Db 2688 MSVADSGEYVCRRANNIDALASIVISVPSAGSPSAPGSMPIRIESSSHVABGTLTD 2747
QY 19 ----- 18
Db 2748 LNCVYVQAHAQVTHKRGSLPSYHQTRGSLRLHVVSPADSGEYVCYVNGSGPLEAS 2807
QY 19 ----- 18
Db 2808 VLVTLIAGSSAVHVPAGAPRIEPPSSRVAEGQTLNLCVYVQAHAQVTHKRG 2867
QY 19 ----- 18
Db 2868 NLPARHQVHPLRLNVPSPADSGEYVSCQVYVNGSGTLEASVLYTIESSPGPIAPGLAQ 2927
QY 19 ----- 20
Db 2928 PIYIASSSHVTEGQTLNLCVYVQAHAQVTVKRGSLPARHQTHGSLRLHVVSPAD 2987
QY 21 ----- 20

```

Db 2988 SGEVYCRAGGPGPEQASFTVTVPPSESSYRLRSPVLSIDPPSTVQGDGDAEFKCL 3047
QY 21 ----- 20
Db 3048 HDGAPISLEKTRNOLEDNHNISPNKSIITIVGTRPSNNGYRCVANSAYVAQSVN 3107
QY 21 ----- 20
Db 3108 LSVHGPFTVSVLPESPVVWVKAVLLECVASAGEBRSARWTRISSTPAKLEQRTYGLMD 3167
QY 21 ----- 20
Db 3168 SHTVLQISSAPSDAGTVVCLANALGTAQKOVEIVDTGMAMPAPQVQAEAEALTVEA 3227
QY 21 ----- 20
Db 3228 GHTATLRCSATGSPARTIHWKRLRSLPFWGHRLEBDTLITPRVAQDSGVYICNATSPAG 3287
QY 21 ----- 20
Db 3288 HAEATILHVESPPYATVTPHEASVQAGEVYVQCLAHGTPPLTFQWSRVGSSLPGHATA 3347
QY 21 ----- 20
Db 3348 RNEELHERRAAPEDSGRTCRVTKVKSAAEAPQOLLVOCPPGSLPANTSTPAGSTPTVQVT 3407
QY 21 ----- 20
Db 3408 POLETKSIGAVEFHCAVPSDRTGLRWFKEGGQLPFGHSVQDGLRIQNLDSQCGTYI 3467
QY 21 ----- 20
Db 3468 CGAHGPMGKAQASQVLYQALPSVLTINITSVQTVVGVHAEFCLALGPKQVWTSKY 3527
QY 21 ----- 20
Db 3528 GGHLPFGTVQSGVYRIAHVELADNGQRCVATNAAGTQSHVLLVQALPOISMPOEVR 3587
QY 21 ----- 20
Db 3588 VPASAAVFPCIASGYPTDISMSKLDGLPPDSRLNNMLPSVOPQADGYVTATN 3647
QY 21 ----- 20
Db 3648 RQGVKAPAHLOVERVVPYTPQPSFLPTIKAVYRKEIKITRPSADGMLLYNG 3707
QY 21 ----- 20
Db 3708 QKRVGSPNTLANRQPDIFSGLVGGRPEFPDAGSGMATIRHPTPLALGNFTVTILRS 3767
QY 21 ----- 20
Db 3768 LTQGLIYGLDAPVNGTSQGRFQGLDINEELYLQGYDYGAIPIRAGLSGRTGCVRELRI 3827
QY 21 ----- 20
Db 3828 QGEIIVFDLNLTAHGISHCPTCDRQPCQNGCHDESSSYVCVPAFTGSRCHESQA 3887
QY 21 ----- 20
Db 3888 LHCHEKMGPDATCVNRPDGNGYTCRCHLGRSGLRCEGVYVTPPSLSGAGSYLALPALT 3947
QY 21 ----- 20
Db 3948 NTHHELRLDVEFKPLADGVLLESGKSGVEDEFSIANYGHLEFAYELSGIAVLRTA 4007
QY 21 ----- 28
Db 4008 EPLALGRHRYSAERLKRKDSGLRVNGRPVLRSSPGKSGGILNHTLLYLQGVESVPLSP 4067
QY 29 ----- 28
Db 4068 ATNKAHPRQCVGEVSVNGKRLDLYSFLSGSGIGQCTYDSSPCERQPCQHGATCMPAGEY 4127

```

```

QY 29 ----- 28
Db 4128 EFQCLCRDGIKGLDCEHENPCQLREPCILHGTCQGTQCLCLPFGSGPQCGSGHIAE 4187
QY 29 ----- 28
Db 4188 SDWHEGSGGNAPRGQYGAVFHDGFLAPGHWFSRLPEVETIELEVRISTASGLIM 4247
QY 29 ----- 28
Db 4248 QGVEVGBAGGKDFISLQDGHLYERYQIGSEARLVEDEPTINDGEMHRYALNRGRG 4307
QY 29 ----- 31
Db 4308 SIQYDEELVSGRSPGPVNAVNAKSGITYIGAPDVATLNGRF 4350

RESULT 15
DYNB.CHIRE
ID DYNB.CHIRE STANDARD; PRT; 4568 AA.
AC 039565;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Dynein beta chain, flagellar outer arm.
GN ODA4 OR ODA-4 OR SUP1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadales; Chlamydomonas.
ON NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2197;
RX MEDLINE=94274778; PubMed=8006077;
RA Mitchell D.R., Brown K.S.;
RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy
chain genes."
RL J. Cell Sci. 107:635-644(1994).
CC -I- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -I- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
CC -I- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
DR EMBL: 002963; AAA19956.1; -
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: PF03028; Dynein_heavy; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
coiled coil.
FT DOMAIN 277 293 COILED COIL (POTENTIAL).
FT DOMAIN 1158 1175 COILED COIL (POTENTIAL).
FT DOMAIN 1372 1400 COILED COIL (POTENTIAL).
FT DOMAIN 1614 1650 COILED COIL (POTENTIAL).
FT DOMAIN 1778 1825 COILED COIL (POTENTIAL).
FT DOMAIN 2017 2045 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 2831 2848 COILED COIL (POTENTIAL).
FT DOMAIN 3106 3162 COILED COIL (POTENTIAL).
FT DOMAIN 3339 3425 COILED COIL (POTENTIAL).
FT DOMAIN 3648 3728 COILED COIL (POTENTIAL).
FT NP_BIND 1919 1926 ATP (POTENTIAL).
FT NP_BIND 2202 2209 ATP (POTENTIAL).
FT NP_BIND 2530 2537 ATP (POTENTIAL).
FT NP_BIND 2879 2886 ATP (POTENTIAL).

```


SQ SEQUENCE 4568 AA; 519961 MM; 9A9A5393C7C36AE7 CRC64;
 Query Match 38.5%; Score 60; DB 1; Length 4568;
 Best Local Similarity 0.8%; Pred. No. 2e+03;
 Matches 19; Conservative 3; Mismatches 6; Indels 2209; Gaps 6;

```

OY 4 HHSHH-----9
Db 1704 HHSHHSEVEVEFEDSCDGPVEVWLVQVNDVSMKLAQVEFRKAIPYDELPRKMIYV 1763
OY 10 -----12
Db 1764 SMOATVVSRTPTFQINAEFDLEEGNEALKVELDQVQQLDIDEINKEQTSIDRK 1823
OY 13 -----12
Db 1824 KLITCTIDVHSRDVOKLIDERVEQKCFQWOSQIARIQSMTKTCQVNICDAEIAVSY 1883
OY 13 -----12
Db 1884 EYIGCGCLITPTDRCFITLQALVLSGAPAGAGTGKTTETWDLARALGIQCYV 1943
OY 13 -----12
Db 1944 NCSQDMYKAGHTYKGLAQIGAWGCFDENRIRIPAVLSVCSQYKYVLAIRAKKERFT 2003
OY 13 -----12
Db 2004 FEDADISLSTVMAETINPGYPGRAELPESLKAIFRPVSVVPDLALICIMLMAEQ 2063
OY 13 -----12
Db 2064 MSKISRKFVILLYKCEULLSKSRHYDKLKAITTLVAGGMRAPDELSEDVLLRAL 2123
OY 13 -----14
Db 2124 RDNIGKLTADDTSTFMGLNDLFPKLELVPRALDKAFDAAMKAATELGYDDQDL 2183
OY 15 -----17
Db 2184 KISHVRELFTVVRMSVFLLAGAGCKTAVWRTLRAONSSGERTYQAVNPRAVTRNELYG 2243
OY 18 -----17
Db 2244 YLHPRAREMKEGLMSVTFRRNANNKTKHQTIVLDGIDAEIESMTVYDNDKMLTAS 2303
OY 18 -----17
Db 2304 NERILPSPMKLLLEINHWYHCSPATVSRGVIFINADVGMQPVASWIDKLEAEYRP 2363
OY 18 -----17
Db 2364 LITLALFTKRYVPCLEHCRNFKTVPLPANOANTICKILEGILPKETVGAPPPDKLL 2423
OY 18 -----17
Db 2424 HYHVFACVNAFGCMLVDKVTYRTQFSKWVSEMKDVGPEKGLVYDYVDEQNCIMV 2483
OY 18 -----17
Db 2484 PWEDRVTKFOYIIGEDTSLVPTVEYETRLTYFLDSLVSNNHYAMPVGNNTGKSALMVNK 2543
OY 18 -----17
Db 2544 LRNDTETMSFYTIINNSLSSEAPALVILMEPLEKSGVYKPGSRKMYFTVDMMKPL 2603
OY 18 -----20
Db 2604 VDKYDTSSIELLRQWVDYHGWDKVIQLEKELINCOMAACNPJAGSFNITPROMGRFY 2663
OY 21 -----20
Db 2664 TFAVQMPAEITRAMYQIIDGHSFDPDVVAKMSKLVDAFCELHRVNHNFPSAVKP 2723
  
```

```

OY 21 -----20
Db 2724 HYQFNLDLSNITQGLTRAJKEYREPVKARLVWHECEHFQDMINEADMAKDFEFV 2783
OY 21 -----20
Db 2784 AVTKFFDDCGMVAIBERPLIYASHASMTYTEDEVYVYALSSVDLRTNLEDKLREYN 2843
OY 21 -----20
Db 2844 ESNAVDLVLEQAMEHVTIRIIDLPRGNAMLVGSGSKSLARLASIIGYEYEQ 2903
OY 21 -----20
Db 2904 SVSSTGINDKENLGLYRKAGTKGPTPTFLMTDQIVNEGFLVYINDLSTGYADLF 2963
OY 21 -----20
Db 2964 TPEDKEAFTNAVREVAAGILDSAEKMDFFLDKVRKFLHYLCSPVGDKERIRAROF 3023
OY 21 -----20
Db 3024 PALVNCMTFDMFGWGEALVSAORELYDVPMEEVYRENIAYHMAVAHQVSEASERP 3083
OY 21 -----20
Db 3084 KEAFRRYNTTPKSYLELISLYKMLQLKRDILRSKERLENGIDKIAQAAAQVTDLQRY 3143
OY 21 -----20
Db 3144 LKEBOIVDEKKAQDELIVSIGEKAIYDQAVEAGREDEAATALQTEVSAPQAECEPD 3203
OY 21 -----20
Db 3204 LLAEPFLIAQAEALSLINKKELSELKSPAPAEIYVAAAQVLYGCGKITPKDRDNA 3263
OY 21 -----20
Db 3264 GKRMADVNSFLSNFTDNDVYVYCEVEKDYISNPGFTPDNINKSACAGLCSRY 3323
OY 21 -----20
Db 3324 INICKFRYQVAPRRAALAEANKKLDVANKKLVIRDEVRLODRVALLEQSLKATE 3383
OY 21 -----20
Db 3384 DKNAIAQADRTPARKQMAERLINGSKNTKGAETKRLSELBGLNGVDVLIASAFVSY 3443
OY 21 -----27
Db 3444 AGPFNNQFRKSLVDEKWLPIIBROIPMTOGLRPLDLDLTDAITKAKMANEGLPDPLSYE 3503
OY 28 -----27
Db 3504 NGAIMSNASRALMDIPOLOGIKWINKETNGLVYIQQSPRYTDVYINCILENGMPLII 3563
OY 28 -----27
Db 3564 ENLPVIDAVLDVYIGKTIKKGRNLIKIGDAEYOYDSRFLYQIKLSNHPHEPVAA 3623
OY 28 -----27
Db 3624 QTLVNVCTEKGLEBOLLAVDHERPDLOQAAGVLSRNLNEYNTIVLENNLLFNIA 3683
OY 28 -----27
Db 3684 NATGNLENIIELEGLIETKTAVEIEKVKLAKOTEHIOIAKAREVYRPVATRSSLYFL 3743
OY 28 -----27
Db 3744 IDNLNLDREVHYSHANFVFLKKGMDMTPGSKDESKVLAERLNGEVDLKRVELLVET 3803
  
```

```

OY 28 ----- 27
Db 3804 TCFVLIGVAGLFRHKLIVATQICNOILRSRGEIHYAKFEYLLRGPKVWGDNPILHM 3863
OY 28 ----- 27
Db 3864 VSDSVWSSVQALKEIDYQGLPEDLIGSSKRRENNELERPEDEPLPGDWKRMQEFDKL 3923
OY 28 -----YGRF 31
          :|||
Db 3924 LFRALRPDRITSAMGRF 3940
    
```

Search completed: May 1, 2003, 14:47:12
 job time : 18 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 10.1719 Seconds

(without alignments)
126.404 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171
Sequence: 1 SRTHSHSEIRTPDINPAWASRGIRPYGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	100.0	87	1	PRRP_HUMAN
2	158	92.4	98	1	PRRP_BOVIN
3	149	87.1	83	1	PRRP_RAT
4	54	31.6	428	1	NER3_HUMAN
5	52	30.4	288	1	Y567_PASUO
6	50.5	29.5	1882	1	Y468_MYCPN
7	50	29.2	428	1	NER3_BOVIN
8	47.5	27.8	345	1	ARQC_BACHD
9	47.5	27.8	533	1	CNC_DROME
10	47	27.5	402	1	EX7L_STROCO
11	47	27.5	676	1	EX1L_HUMAN
12	46.5	27.2	240	1	PLSC_HELPY
13	46.5	27.2	1693	1	POLN_HEVU
14	46.5	27.2	1693	1	POLN_HEVU
15	46.5	27.2	1783	1	Y468_MYCCE
16	46	26.9	342	1	Y762_METTA
17	46	26.9	347	1	Y576_METTA
18	46	26.9	383	1	CYCR_CHRYI
19	46	26.9	1400	1	RIF1_SCHPO
20	45.5	26.6	239	1	RIPL_SYNY3
21	45.5	26.6	407	1	Y116_MYCUP
22	45	26.3	213	1	SRN2_YEAST
23	45	26.3	350	1	HNM2_DROME
24	45	26.3	510	1	YCOB_ECOLI
25	45	26.3	798	1	UNR_RAT
26	45	26.3	940	1	UVRA_VIBCH
27	45	26.3	962	1	UVRA_METTH
28	45	26.3	973	1	UVRA_METTH
29	45	26.3	973	1	UVRA_METTH
30	44.5	26.0	860	1	VG12_BPBO3
31	44	25.7	137	1	SMR2_RAT
32	44	25.7	364	1	YAIW_ECOLI
33	44	25.7	386	1	CRTY_AGRAU

SEA ID NO: 61
Database: SwissProt-40
AC NO: P81277

34	44	25.7	476	1	YAIW_ECOLI	P30143 escherichia
35	44	25.7	581	1	POL_MLYRK	P31795 radiation m
36	44	25.7	591	1	PYRG_HUMAN	P17812 homo sapien
37	44	25.7	719	1	NRP1_YEAST	P32770 saccharomyc
38	44	25.7	843	1	POL_MLYRK	P03357 akt murine
39	44	25.7	1087	1	RP16_HUMAN	P03357 akt murine
40	44	25.7	1087	1	RP16_MOUSE	P03357 akt murine
41	44	25.7	1196	1	POL_MLYRK	P03357 akt murine
42	44	25.7	1196	1	POL_MLYRK	P03357 akt murine
43	43.5	25.4	568	1	RSL_RICPR	P11227 radiation m
44	43.5	25.4	572	1	GAG_IPHA	P04023 hamster int
45	43.5	25.4	770	1	AVP3_ARATH	P31414 arabidopsis

ALIGNMENTS

RESULT 1	PRRP_HUMAN	STANDARD;	PRT;	87 AA.
ID	P81277;			
AC	P81277;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Proactin-releasing peptide precursor (PRRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRRP31; Proactin-releasing peptide PRRP20].			
GN	PRH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE:Brain;			
RA	MEDLINE:98268781; PubMed:9607765;			
RA	Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,			
RA	Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,			
RA	Kurokawa T., Nishimura O., Onda H., Fujino M.;			
RT	"A proactin-releasing peptide in the brain.";			
RL	Nature 393:272-276(1998).			
RN	[2]			
RP	TISSUE SPECIFICITY.			
RA	MEDLINE:99426552; PubMed:10498338;			
RA	Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S.,			
RA	Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H.,			
RT	"Tissue distribution of proactin-releasing peptide (PRRP) and its			
RL	receptor.";			
RL	Regul. Pept. 83:1-10(1999).			
CC	-I- FUNCTION: Stimulates proactin (PRT) release and regulates the			
CC	expression of proactin through its receptor GPR10. May stimulate			
CC	lactotrophs directly to secrete prl.			
CC	-I- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOPHALANUS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: AB015419; BAA29027.1;			
DR	MI: 602663;			
KW	Hormone; Amidation; Signal.			
FT	SIGNAL			
FT	PEPTIDE			
FT	PEPTIDE			
FT	MOD_RSS			
FT	SEQUENCE			
Query Match	87 AA;	9639 MM;	229A2F350CF981B CRK64;	
	100.0%;	Score 171;	DB 1;	Length 87;


```

RT      Cloning expression, and chromosomal mapping of a human ganglioside
RT      sialidase." ;
RL      Biochem. Biophys. Res. Commun. 261:21-27(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE-Skeletal muscle;
RX      PubMed-10861246;
RA      Monti E., Bassi M.T., Papini N., Riboni M., Manzoni M., Venerando B.,
RA      Croci G., Preti A., Ballabio A., Tettamanzi G., Borsani G.;
RT      Identification and expression of NEU3, a novel human sialidase
RT      associated to the plasma membrane." ;
RL      Biochem. J. 349:343-351(2000).
CC      -1- FUNCTION: Plays a role in modulating the ganglioside content of
CC      the lipid bilayer at the level of membrane-bound sialyl
CC      glycoconjugates.
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC      alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC      oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC      synthetic substrates.
CC      -1- SUBCELLULAR LOCATION: Membrane-associated.
CC      -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle, testis,
CC      adrenal gland and thymus, followed by pancreas, liver, heart and
CC      thymus. Weakly expressed in kidney, placenta, brain and lung.
CC      -1- MISCELLANEOUS: Optimum pH is 3.8.
CC      -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC      -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/annoucnce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AB008185; BAA82611.1; -
DR      EMBL; Y18563; CAB96131.1; ALT_INTL.
DR      Genew; HGNC:7760; NEU3.
DR      MIM; 604617; -
DR      InterPro; IPR002860; GH_BNR.
DR      Pfam; PF02012; BNR; 6.
KW      Hydrolase; Glycosidase; Membrane; Repeat.
FT      REPEAT 129 140 BNR 1.
FT      REPEAT 203 214 BNR 2.
FT      REPEAT 254 265 BNR 3.
FT      SITE 24 27 FRIP MOTIF.
FT      ACT_SITE 25 25 BY SIMILARITY.
FT      ACT_SITE 45 45 POTENTIAL.
FT      ACT_SITE 50 50 POTENTIAL.
FT      ACT_SITE 87 87 POTENTIAL.
FT      ACT_SITE 225 225 POTENTIAL.
FT      ACT_SITE 245 245 POTENTIAL.
FT      ACT_SITE 340 340 BY SIMILARITY.
FT      ACT_SITE 370 370 POTENTIAL.
FT      ACT_SITE 387 387 POTENTIAL.
SQ      SEQUENCE 428 AA; 48252 MW; 35D1DD9359A78C98 CRC64;

Query Match 31.6%; Score 54; DB 1; Length 428;
Best Local Similarity 37.0%; Pred. No. 1.5;
Matches 10; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

CY 2 RTHRSMLRPPDINPANTASGIRPV 28
DB 195 RTRPSLMIYSDDLGVTHNGRLIRP 221

RESULT 5
ID Y587_PASMD STANDARD; PRT; 288 AA.
AC Q9CNS6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

01 Hypothetical protein PM0587.
 02
 03 *Pasteurella multocida*.
 04
 05 *Bacteria*; *Proteobacteria*; *gamma* subdivision; *Pasteurellaceae*;
 06 *Pasteurella*.
 07
 08 NCBI_TaxID=747;
 09
 10 [1]
 11
 12 SEQUENCE FROM N.A.
 13
 14 STRAIN=Pm70;
 15 MEDLINE=21145866; Pubmed=11248100;
 16
 17 May B.T., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 18
 19 *Complete genomic sequence of *Pasteurella multocida* Pm70.";
 20 Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 21
 22 -1- SIMILARITY: BELONGS TO THE PUCOSANINE KINASE FAMILY.
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 49

```

Query March 30.4% Score 52; DB 1; Length 288;
Best Local Similarity 37.5%; Pred. No. 1.9;
Matches 9; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

CY 5 RNSMERTPDINPAPVASRGIRPV 28
    ||| |::: |||
Db 20 NKKEKIHGEMHEAWIIDGDIQPV 43

```

```

RESULT 6
Y468_MYCPN
ID Y468_MYCPN STANDARD: PRT: 1882 AA.
AC P75109; Q50317;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein MG468 homolog (R05_cri1882).
KW MP0684 OR MP158.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129.
RX MEDLINE=97105885; Pubmed=8948633;
RA Himmelfreich R., Hilbert N., Plegens H., Pirk E., Li B.-C.,
  Hermann R.;
RT *Complete sequence analysis of the genome of the bacterium Mycoplasma
  pneumoniae.;
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP SEQUENCE OF 1-1848 FROM N.A.
RC STRAIN=ATCC 29342 / M129.
RX MEDLINE=96177562; Pubmed=8664303;
RA Hilbert H., Himmelfreich R., Plegens H., Hermann R.;
RT *Sequence analysis of 56 kb from the genome of the bacterium
  Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a
  cluster of ribosomal protein genes.";
RL Nucleic Acids Res. 24:628-639(1996).
RN [3]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=21088919; Pubmed=11271496;
RA Regula J.T., Ueberle B., Boguth G., Georg A., Schnoelzer M.,
  Hermann R., Frank R.;
RT *Towards a two-dimensional proteome map of Mycoplasma pneumoniae.";

```


DR Pfam: PF01118; Semaldhyde_dh; 1.
 DR Pfam: PF02774; Semaldhyde_dhc; 1.
 DR Prodom: PD003765; AgPR_act_site; 1.
 DR PROSITE: PS01224; ARCC; 1.
 KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT ACT SITE 149 149 BY SIMILARITY.
 SQ SEQUENCE 345 AA; 38188 MW; 3E9F45DD09FC68EA CRC64;

Query Match 27.8%; Score 47.5; DB 1; Length 345;
 Best Local Similarity 45.5%; Pred. No. 11;
 Matches 10; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 8 MEIRTPDINPARYASGIRVVG 29
 Db 101 LRINEPDYERAY-KRQAAVVG 121

RESULT 9

CNC_DROME STANDARD; PRT: 533 AA.

AC P20482;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Segmentation protein cap'n/collar.
 GN CNC.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.
 RX MEDLINE-92001535; PubMed-1911393;
 RA Mohler J., Van K., Leung S., Epstein A.;
 RT Segmentally restricted, cephalic expression of a leucine zipper gene
 during Drosophila embryogenesis";
 Mech. Dev. 34:3-9(1991).

CC -1- FUNCTION: PLAYS A ROLE IN CEPHALIC PATTERNING. PROBABLE SUBUNIT
 OF A HETEROODIMERIC REGULATORY PROTEIN INVOLVED IN THE CONTROL OF
 HEAD MORPHOGENESIS.

CC -1- DEVELOPMENTAL STAGE: LOCALIZED TO THE MANDIBULAR SEGMENT AND THE
 HYPOPHARYNGEAL AND LABRAL PRIMORDIA FIRST DETECTABLE IN LATE
 BLASTODERM STAGES.

CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY, CNC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib.ch).

DR EMBL: M37495; AA859246.1; -

DR PIR: A33111; A33111.

DR HSSP: P34707; ISKN.

DR TRANSFAC: T01988; -

DR FLYBASE: FBgn0000338; CNC.

DR InterPro: IPR004827; TF_bzip.

DR Pfam: PR00170; bzip; 1.

DR SMART: SM00336; BRLZ; 1.

DR PROSITE: PS00036; BZIP_BASIC; 1.

KW Transcription regulation; Activator; DNA-binding; Nuclear protein;

KW Developmental protein.

FT DNA_BIND 351 369 BASIC MOTIF.

FT DOMAIN 387 408 LEUCINE-ZIPPER.

FT DOMAIN 497 529 GIN-RICH.

FT CONFLICT 349 349 I -> L (IN REF. 1; AA859246).

SQ SEQUENCE 533 AA; 56948 MW; EADF9A5D6CASC5F CRC64;

Query Match 27.8%; Score 47.5; DB 1; Length 533;
 Best Local Similarity 37.5%; Pred. No. 18;

Matches 9; Conservative 5; Mismatches 9; Indels 1; Gaps 1;
 QY 4 HRSME-IRTPDINPARYASGIR 26
 Db 409 HHVGFQYLRDPSGNCWPADYSIQ 432

RESULT 10

EXTL_STRCO STANDARD; PRT: 402 AA.

AC Q9PBM3; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)

DE (Exonuclease VII large subunit).

GN XSEA OR SC05056 OR SCK7.29c.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteriae; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

NCBI_TaxID=1902;

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2) / M145;

RX MEDLINE-21996410; PubMed-1200953;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT *Complete genome sequence of the model actinomycete Streptomyces

coelicolor A3(2).";

RI Nature 417:141-147(2002).

CC -1- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE

ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER

INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: EXONUCLEOLYTIC cleavage in either 5'- to 3'-

OR 3'- to 5'-direction to yield nucleoside 5'-phosphates.

CC -1- SUBUNIT: HETERODIMER COMPOSED OF LARGE AND SMALL SUBUNITS (BY

SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: BELONGS TO THE XSEA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib.ch).

DR EMBL: AL391754; CAC05901.1; -

DR InterPro: IPR003753; Exonuc_VII_L.

DR InterPro: IPR004365; tRNA_antl.

DR Pfam: PF01336; tRNA_antl; 1.

DR Pfam: PF02601; Exonuc_VII_L; 1.

DR TIGRFRAMS: TIGR00237; xsea; 1.

KW Hydroxylase; Nuclease; Exonuclease; Complete proteome.

SQ SEQUENCE 402 AA; 43882 MW; 1A5929A372B4E08 CRC64;

Query Match 27.5%; Score 47; DB 1; Length 402;
 Best Local Similarity 47.6%; Pred. No. 15;
 Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

QY 17 PAWYASRG-----IRPVG 29
 Db 89 PEMYAPRGQLSLRAAIKIPVG 109

OY 1 SRRHSMETRPDIN-PANV 20
DB 197 ARRLVMEESTYDPDENSEPTW 217

RESULT 13

ID POLN_HEVBU STANDARD; PRT: 1693 AA.

AC P29324:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Non-structural polypeptide [Contains: RNA-directed RNA polymerase
(EC 2.7.7.48); Helicase].
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID-31767;
RN (1)

RP SEQUENCE FROM N.A.
RX MEDLINE-92024067; PubMed-1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R.:
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome";
RT Virology 185:120-131(1991).
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
(RNA)(N).

CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).

CC EMBL: W73218; AAA5734.1; -
DR PIR: A40778; KMWME.

DR MEROPS: C41.001; -
DR InterPro: IPR002589; Alpp.
DR InterPro: IPR001230; Prenyl_site.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR Pfam: PF01661; Alpp; 1.
DR SMART: SM00506; Alpp; 1.
KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP BIND
SQ SEQUENCE 1693 AA: 185191 MW: 2F355E46E96D219B CRC64:

Query Match 27.2%; Score 46.5; DB 1: Length 1693;
Best Local Similarity 37.0%; Pred. No. 89;
Matches 10; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

OY 2 RTRHSMETRPDIN-PANV-ASRGTRP 27
DB 904 RNRHSGDELYLPFLAARWEANRPTRP 930

RESULT 14

ID POLN_NEVNY STANDARD; PRT: 1693 AA.

AC 004610;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Non-structural polypeptide [Contains: RNA-directed RNA polymerase
DE (EC 2.7.7.48); Helicase].
OS Hepatitis E virus (strain Myanmar) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID-31769;
RN (1)

RP SEQUENCE FROM N.A.
RX MEDLINE-93227573; PubMed-8470371;
RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,
RA Rikihisa T., Wilm K.:
RT "Sequence and gene structure of the hepatitis E virus isolated from
RT Myanmar";
RT Virology 7:95-109(1993).

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
(RNA)(N).

CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).

CC EMBL: D10330; BA01172.1; -
DR MEROPS: C41.001; -

DR InterPro: IPR002589; Alpp.
DR InterPro: IPR001230; Prenyl_site.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR Pfam: PF01661; Alpp; 1.
DR SMART: SM00506; Alpp; 1.
KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP BIND
SQ SEQUENCE 1693 AA: 185215 MW: AAB4C9140A7E21EA CRC64:

Query Match 27.2%; Score 46.5; DB 1: Length 1693;
Best Local Similarity 37.0%; Pred. No. 89;
Matches 10; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

OY 2 RTRHSMETRPDIN-PANV-ASRGTRP 27
DB 904 RNRHSGDELYLPFLAARWEANRPTRP 930

RESULT 15

ID Y468_MYCE STANDARD; PRT: 1783 AA.

AC Q49460;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M6468.
GN M6468.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID-2097;
RN (1)

RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 33530 / G-37;
RX MEDLINE-96026346; PubMed-7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,

RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. II, Venter J.C.;
 RT "The minimal gene complement of *Mycoplasma genitalium*."; *Science* 270:397-403(1995).
 RL [2]
 RN REVISIONS.
 RP Fraser C.M., Cocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. II, Venter J.C.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 879-985 FROM N.A.
 RC SRRAIN-ATCC 33530 / G-37;
 RX MEDLINE-94075230; PubMed-8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III,
 RT "A survey of the *Mycoplasma genitalium* genome by using random
 RT sequencing."; *J. Bacteriol.* 175:7918-7930(1993).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: SOME, TO MG064.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39728; AAC72488.1; -
 DR EMBL: U01808; AAD12339.1; -
 DR TIGR: MG668; -
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 917 937 POTENTIAL.
 FT TRANSMEM 967 987 POTENTIAL.
 FT TRANSMEM 1010 1030 POTENTIAL.
 FT TRANSMEM 1084 1104 POTENTIAL.
 FT TRANSMEM 1660 1680 POTENTIAL.
 FT TRANSMEM 1709 1729 POTENTIAL.
 FT TRANSMEM 1730 1750 POTENTIAL.
 FT TRANSMEM 1752 1772 POTENTIAL.
 SQ SEQUENCE 1783 AA; 200168 MW: 87BD575AEC2E374B CRC64,
 Query Match 27.28; Score 46.5; DB 1; Length 1783;
 Best Local Similarity 29.08; Pred. No. 94;
 Matches 9; Conservative 6; Mismatches 13; Indels 3; Gaps 1,
 QY 1 SRTGRHSMEITPDINPMATYASRGIRPVGRF 31
 Db 1118 NRNPYKINLCOTPEPQSGWYA---IQPYSNF 1145

Search completed: May 1, 2003, 14:33:19
 Job time : 13.1719 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:13 ; Search time 15.0156 seconds
(without alignments)
60,744 Million cell updates/sec

Title: US-09-446-543a-61
Perfect score: 171
Sequence: 1 SRTIRHSMETRPDINPAMWASRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCRU5.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the total being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	100.0	31	3	US-09-105-678A-9
2	171	100.0	31	3	US-09-105-678A-43
3	171	100.0	31	4	US-08-776-971-61
4	171	100.0	31	4	US-09-421-208-9
5	171	100.0	31	4	US-09-421-208-9
6	171	100.0	31	4	US-09-560-915-13
7	171	100.0	32	3	US-09-105-678A-44
8	171	100.0	32	4	US-08-776-971-62
9	171	100.0	32	4	US-09-421-208-44
10	171	100.0	33	3	US-09-105-678A-45
11	171	100.0	33	4	US-08-776-971-63
12	171	100.0	33	4	US-09-421-208-45
13	171	100.0	87	4	US-08-776-971-59
14	171	100.0	87	4	US-08-776-971-135
15	171	100.0	87	4	US-08-776-971-138
16	158	92.4	31	3	US-09-105-678A-7
17	158	92.4	31	3	US-09-105-678A-31
18	158	92.4	31	4	US-08-776-971-5
19	158	92.4	31	4	US-08-776-971-97
20	158	92.4	31	4	US-09-421-208-7
21	158	92.4	31	4	US-09-421-208-31
22	158	92.4	31	4	US-09-560-915-13
23	158	92.4	32	3	US-09-105-678A-32
24	158	92.4	32	4	US-08-776-971-6
25	158	92.4	32	4	US-09-421-208-32
26	158	92.4	33	3	US-09-105-678A-33
27	158	92.4	33	4	US-08-776-971-7

28	158	92.4	33	4	US-09-421-208-33	Sequence 33, App1
29	158	92.4	98	4	US-08-776-971-1	Sequence 1, App1
30	158	92.4	98	4	US-08-776-971-44	Sequence 44, App1
31	158	92.4	98	4	US-08-776-971-122	Sequence 122, App1
32	158	92.4	98	4	US-08-776-971-131	Sequence 131, App1
33	158	92.4	98	4	US-08-776-971-136	Sequence 136, App1
34	154	90.1	98	4	US-08-776-971-115	Sequence 115, App1
35	149	87.1	31	3	US-09-105-678A-8	Sequence 8, App1
36	149	87.1	31	3	US-09-105-678A-37	Sequence 37, App1
37	149	87.1	31	4	US-09-172-353-4	Sequence 4, App1
38	149	87.1	31	4	US-08-776-971-47	Sequence 47, App1
39	149	87.1	31	4	US-09-421-208-8	Sequence 8, App1
40	149	87.1	31	4	US-09-421-208-37	Sequence 37, App1
41	149	87.1	31	4	US-09-560-915-14	Sequence 14, App1
42	149	87.1	32	3	US-09-105-678A-38	Sequence 38, App1
43	149	87.1	32	4	US-08-776-971-48	Sequence 48, App1
44	149	87.1	32	4	US-09-421-208-38	Sequence 38, App1
45	149	87.1	32	4	US-09-421-208-38	Sequence 38, App1

ALIGNMENTS

RESULT 1
US-09-105-678A-9
Sequence 9, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suemaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-9

Query Match 100.0%; Score 171, DB 3; Length 31,
Best Local Similarity 100.0%; Pred. No. 7, le-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

09 1 SRTIRHSMETRPDINPAMWASRGIRPVGRF 31

Db 1 SRTSRHSMETPTDINPAMYASRGIRPVGRF 31

US-09-105-678A-43
Sequence 43, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Motiwa, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conllo, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-43

Query Match 100.0%; Score 171; DB 3; Length 31;

Best Local Similarity 100.0%; Pred. No. 7,1e-19;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTSRHSMETPTDINPAMYASRGIRPVGRF 31

Db 1 SRTSRHSMETPTDINPAMYASRGIRPVGRF 31

RESULT 3

US-08-776-971-61

Sequence 61, Application US/08776971B

Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Motiwa, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conllo, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 61:

US-08-776-971-61

Query Match 100.0%; Score 171; DB 4; Length 31;

Best Local Similarity 100.0%; Pred. No. 7,1e-19;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTSRHSMETPTDINPAMYASRGIRPVGRF 31

Db 1 SRTSRHSMETPTDINPAMYASRGIRPVGRF 31

RESULT 4

US-09-421-208-9

Sequence 9, Application US/09421208

Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Motiwa, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-9

Query Match 100.0%; Score 171; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 7,1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 5
US-09-421-208-43
Sequence 43, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-43

Query Match 100.0%; Score 171; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 7,1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 6
US-09-560-915-15
Sequence 15, Application US/09560915
Patent No. 6383764
GENERAL INFORMATION:
APPLICANT: Civeill, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Therapeutic Compositions and Methods
TITLE OF INVENTION: Relating to Prolactin Releasing Peptide (Pr-P)
FILE REFERENCE: P-UC 3534
CURRENT APPLICATION NUMBER: US/09/560,915
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 31
TYPE: Prt
ORGANISM: Homo Sapien
US-09-560-915-15

Query Match 100.0%; Score 171; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 7,1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 7
US-09-105-678A-44
Sequence 44, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A

FILED DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-44

Query Match 100.0%; Score 171; DB 3; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.3e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTNRSMETRTPDINPAMYASRGIRPVGRF 31
DB 1 SRTNRSMETRTPDINPAMYASRGIRPVGRF 31

RESULT 8

US-08-776-971-62
Sequence 62, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Nishimura, Shuji
Nabata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Kikushima, Shoji
Kilada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FRAGMENT TYPE: Internal
SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-08-776-971-62

Query Match 100.0%; Score 171; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.3e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTNRSMETRTPDINPAMYASRGIRPVGRF 31
DB 1 SRTNRSMETRTPDINPAMYASRGIRPVGRF 31

RESULT 9

US-09-421-208-44
Sequence 44, Application US/09421208
Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-44

Query Match 100.0%; Score 171; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.3e-19;

APPLICANT: Ninuma, Shuji
Nabata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Fukusumi, Shoichi
Kitada, Chieko

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 192T LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: linear
MOLECULE TYPE: peptide
US-09-421-208-45

Query Match 100.0%; Score 171; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 7,6e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETPTDINPMTASRGIRPVGRF 31
DB 1 SRTHRSMETPTDINPMTASRGIRPVGRF 31

RESULT 13
US-08-776-971-59
Sequence 59, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-FEB-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47716
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 59:
US-08-776-971-59

Query Match 100.0%; Score 171; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETPTDINPMTASRGIRPVGRF 31
DB 23 SRTHRSMETPTDINPMTASRGIRPVGRF 53

RESULT 14
US-08-776-971-135
Sequence 135, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-FEB-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G,
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 135:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 135:
US-08-776-971-135

Query Match 100.0%; Score 171; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 2,4e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTSRHSMETPTDINPWTASRGIRPVGRF 31
DB 23 SRTSRHSMETPTDINPWTASRGIRPVGRF 53

RESULT 15
US-08-776-971-138

Sequence 138, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 138:
US-08-776-971-138

Query Match 100.0%; Score 171; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 2,4e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTSRHSMETPTDINPWTASRGIRPVGRF 31
DB 23 SRTSRHSMETPTDINPWTASRGIRPVGRF 53

Search completed: May 1, 2003, 14:33:54
Job time: 16.0156 secs

100

100

100

100

100

GenCore version 5.1.4.D5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:46:12 ; Search time 15 Seconds
(without alignments)
64.730 Million cell updates/sec

Title: SEQ74-PLUS-73
Sequence: 1 SRHXHSMEXRTPDINPAMWYXXRGIRVGRXX 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	31	3	US-09-105-678A-8
2	147	94.2	31	3	US-09-105-678A-9
3	147	94.2	31	3	US-09-105-678A-37
4	147	94.2	31	3	US-09-105-678A-43
5	147	94.2	31	4	US-09-172-353-4
6	147	94.2	31	4	US-08-776-971-47
7	147	94.2	31	4	US-08-776-971-61
8	147	94.2	31	4	US-09-421-208-8
9	147	94.2	31	4	US-09-421-208-9
10	147	94.2	31	4	US-09-421-208-37
11	147	94.2	31	4	US-09-421-208-43
12	147	94.2	31	4	US-09-560-915-14
13	147	94.2	31	4	US-09-560-915-15
14	147	94.2	32	3	US-09-105-678A-38
15	147	94.2	32	3	US-09-105-678A-44
16	147	94.2	32	4	US-08-776-971-48
17	147	94.2	32	4	US-08-776-971-62
18	147	94.2	32	4	US-09-421-208-38
19	147	94.2	32	4	US-09-421-208-44
20	147	94.2	33	3	US-09-105-678A-39
21	147	94.2	33	3	US-09-105-678A-45
22	147	94.2	33	4	US-08-776-971-49
23	147	94.2	33	4	US-08-776-971-63
24	147	94.2	33	4	US-09-421-208-39
25	147	94.2	33	4	US-09-421-208-45
26	147	94.2	83	4	US-08-776-971-45
27	147	94.2	83	4	US-08-776-971-124

28	147	94.2	83	4	US-08-776-971-137	Sequence 137, App
29	147	94.2	87	4	US-08-776-971-59	Sequence 59, Appl
30	147	94.2	87	4	US-08-776-971-135	Sequence 135, App
31	147	94.2	87	4	US-08-776-971-138	Sequence 138, App
32	146	93.6	31	3	US-09-105-678A-7	Sequence 7, Appl
33	146	93.6	31	3	US-09-105-678A-31	Sequence 31, Appl
34	146	93.6	31	3	US-08-776-971-5	Sequence 5, Appl
35	146	93.6	31	4	US-08-776-971-97	Sequence 97, Appl
36	146	93.6	31	4	US-09-421-208-7	Sequence 7, Appl
37	146	93.6	31	4	US-09-421-208-31	Sequence 31, Appl
38	146	93.6	31	4	US-09-560-915-13	Sequence 13, Appl
39	146	93.6	32	3	US-08-105-678A-32	Sequence 32, Appl
40	146	93.6	32	4	US-08-776-971-6	Sequence 6, Appl
41	146	93.6	32	4	US-09-421-208-32	Sequence 32, Appl
42	146	93.6	33	3	US-09-105-678A-33	Sequence 33, Appl
43	146	93.6	33	4	US-08-776-971-7	Sequence 7, Appl
44	146	93.6	33	4	US-09-421-208-33	Sequence 33, Appl
45	146	93.6	98	4	US-08-776-971-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-105-678A-8
Sequence 8, Application US-09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-09/105,678A
FILING DATE: 26-JUN-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-8
Query Match 94.2% ; Score 147 ; DB 3 ; Length 31 ;
Best Local Similarity 83.9% ; Pred. No. 1.2e-12 ;
Matches 26 ; Conservative 0 ; Mismatches 5 ; Indels 0 ; Gaps 0 ;

QY 1 SRHXHSMEXRTPDINPAMWYXXRGIRVGRP 31

Db 1 SRHHSHMETRTDINPAMVYTGIRPVGRF 31

RESULT 2

US-09-105-678A-9
Sequence 9, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street

CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-09-105-678A-9

Query Match 94.2%; Score 147; DB 3; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRHHSHMETRTDINPAMVYTGIRPVGRF 31

Db 1 SRHHSHMETRTDINPAMVYTGIRPVGRF 31

RESULT 3

US-09-105-678A-37
Sequence 37, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:
APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street

CITY: Boston

STATE: MA
COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-09-105-678A-37

Query Match 94.2%; Score 147; DB 3; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRHHSHMETRTDINPAMVYTGIRPVGRF 31

Db 1 SRHHSHMETRTDINPAMVYTGIRPVGRF 31

RESULT 4

US-09-105-678A-43
Sequence 43, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:
APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street

CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-43

Query Match 94.2%; Score 147; DB 3; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

RESULT 5
US-09-172-353-4
Sequence 4, Application US/09172353
Patent No. 6197530
GENERAL INFORMATION:
APPLICANT: Stricker-Kongra, Alain
APPLICANT: Gu, Wei
TITLE OF INVENTION: GPRIO AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS
FILE REFERENCE: 07334/102001
CURRENT APPLICATION NUMBER: US/09/172,353
CURRENT FILING DATE: 1998-10-14
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Mus musculus
US-09-172-353-4

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

RESULT 6
US-08-776-971-47
Sequence 47, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hlunuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776, 971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-08-776-971-47

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

RESULT 7
US-08-776-971-61
Sequence 61, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hlunuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776, 971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-08-776-971-61

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHSMEKRTPDINPANYXXNGIRPVGRF 31
Db 1 SRTFRSMETRTPDINPANYASRGIRPVGRF 31

RESULT 8
US-09-421-208-8
Sequence 8, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-8

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHSMEKRTPDINPANYXXNGIRPVGRF 31
Db 1 SRAHSHMETRTPDINPANYTGKIRPVGRF 31

RESULT 9
US-09-421-208-9
Sequence 9, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-9

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHSMEKRTPDINPANYXXNGIRPVGRF 31

Db 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31

RESULT 10

US-09-421-208-37
Sequence 37, Application US/09421208
Patent No. 628561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-37

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31
Db 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31

RESULT 11

US-09-421-208-43
Sequence 43, Application US/09421208
Patent No. 628561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-43

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31
Db 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31

RESULT 12

US-09-560-915-14
Sequence 14, Application US/09560915
Patent No. 6383764
GENERAL INFORMATION:
APPLICANT: Cleveli, Oliver
APPLICANT: Lin, Steven
TITLE OF INVENTION: Therapeutic Compositions and Methods
RELATING TO Prolactin Releasing Peptide (PRP)
FILE REFERENCE: P-UC 3534
CURRENT APPLICATION NUMBER: US/09/560,915
FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 31
TYPE: PRP
ORGANISM: Rattus
US-09-560-915-14

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31
Db 1 SRTHSHMETRTPDINPAMYASRGIRPVGRF 31

RESULT 13

US-09-560-915-15
; Sequence 15, Application US/09560915
; Patent No. 6383764
; GENERAL INFORMATION:
; APPLICANT: Clevelin, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Inherapeutic Compositions and Methods
; FILE REFERENCE: P-UC 3534
; CURRENT APPLICATION NUMBER: US/09/560,915
; CURRENT FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-560-915-15

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHSMEKRPDINPAMVYXRGIRPVGRF 31
DB 1 SRTNRHSMETRPDINPAMVYXRGIRPVGRF 31

RESULT 14
US-09-105-678A-38
; Sequence 38, Application US/09105678A
; Patent No. 6103882
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,678A
; FILING DATE: 26-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 172118/1997
; FILING DATE: 27-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 48466-342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-105-678A-38

Query Match 94.2%; Score 147; DB 3; Length 32;
Best Local Similarity 83.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHSMEKRPDINPAMVYXRGIRPVGRF 31
DB 1 SRTNRHSMETRPDINPAMVYXRGIRPVGRF 31

RESULT 15
US-09-105-678A-44
; Sequence 44, Application US/09105678A
; Patent No. 6103882
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,678A
; FILING DATE: 26-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 172118/1997
; FILING DATE: 27-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 48466-342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-105-678A-44

Query Match 94.2%; Score 147; DB 3; Length 32;
Best Local Similarity 83.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHSMEKRPDINPAMVYXRGIRPVGRF 31
DB 1 SRTNRHSMETRPDINPAMVYXRGIRPVGRF 31

Search completed: May 1, 2003, 14:49:48
Job time: 15 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 12.0312 Seconds

(without alignments)
188.386 Million cell updates/sec

Title: US-09-446-543a-74

Perfect score: 47

Sequence: 1 SRXHXSMEXR 11

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Database : SPTREMBL.21.*

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rentent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriophage:*
17: sp.archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	89.4	98	6	Q8WN12
2	35	74.5	450	4	Q95367
3	35	74.5	451	4	Q12804
4	34	72.3	672	16	Q8XBC4
5	33	70.2	340	2	Q9RHI9
6	33	70.2	4133	13	Q9DE12
7	33	70.2	4134	13	Q8QGX4
8	32	68.1	196	17	Q9BNP4
9	32	68.1	277	11	Q9D8A2
10	32	68.1	286	11	Q9DA55
11	32	68.1	383	16	Q84156
12	32	68.1	450	11	Q9DA19
13	32	68.1	648	16	Q9XOW5
14	32	68.1	712	16	Q8Y280
15	31	66.0	141	6	Q8WMO1
16	31	66.0	165	2	Q9JN22

17	31	66.0	685	5	Q8T016	Q8T016 drosophila
18	31	66.0	756	5	Q9U116	Q9U116 drosophila
19	31	66.0	800	5	Q9V731	Q9V731 drosophila
20	31	66.0	905	5	Q9W1R8	Q9W1R8 drosophila
21	30	63.8	147	5	Q9B168	Q9B168 blackburnia
22	30	63.8	274	10	Q949G1	Q949G1 oryza sativ
23	30	63.8	330	16	Q9K7F3	Q9K7F3 bacillus ha
24	30	63.8	330	16	Q9A0B5	Q9A0B5 streptococc
25	30	63.8	332	16	Q05268	Q05268 bacillus su
26	30	63.8	387	16	Q8XXK0	Q8XXK0 raietonia s
27	30	63.8	393	5	Q23359	Q23359 caenorhabdi
28	30	63.8	394	10	Q9AS48	Q9AS48 oryza sativ
29	30	63.8	462	11	Q9CT07	Q9CT07 mus musculu
30	30	63.8	509	16	Q9XA23	Q9XA23 streptomyce
31	30	63.8	528	3	Q9C2S0	Q9C2S0 neurospora
32	30	63.8	599	16	Q9KJ40	Q9KJ40 vibrio chol
33	30	63.8	707	10	Q9ATSO	Q9ATSO oryza sativ
34	30	63.8	1889	10	Q94H07	Q94H07 oryza sativ
35	29	61.7	96	8	Q94VZ0	Q94VZ0 pyralisella l
36	29	61.7	122	16	Q9K0W8	Q9K0W8 neisseria m
37	29	61.7	122	16	Q9J710	Q9J710 neisseria m
38	29	61.7	147	17	Q8TPV9	Q8TPV9 methanosarc
39	29	61.7	152	10	Q8S5H7	Q8S5H7 oryza sativ
40	29	61.7	150	11	Q91VZ5	Q91VZ5 mus musculu
41	29	61.7	177	11	Q8Q2T3	Q8Q2T3 mus musculu
42	29	61.7	194	10	Q41320	Q41320 secale cere
43	29	61.7	220	12	Q9Q3U4	Q9Q3U4 cymbidium m
44	29	61.7	233	3	Q9P826	Q9P826 candida alb
45	29	61.7	336	10	Q9AV29	Q9AV29 oryza sativ

ALIGNMENTS

RESULT 1

ID	Q8WN12	PRELIMINARY;	PRT;	98 AA.
AC	Q8WN12			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DR	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Proprolactin-releasing peptide.			
OS	Ovis aries (sheep).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Caprinae; Ovis.			
OX	NCBI_TaxID=9940;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Curlewis J.D., Kusters D.H.L., Barclay J.L., Anderson S.T.;			
RT	*Proactin-releasing peptide (PrRP) in the ewe: cDNA cloning, mRNA			
RT	distribution and effects on prolactin secretion in vitro and in			
RT	vivo.*			
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF450453; AAL47178.1;			
SO	SEQUENCE 98 AA; 10513 MW; 2A5331ED62CAAD5 CRC64;			

Query Match	89.4%	Score 42;	DB 6;	Length 98;
Best Local Similarity	72.7%	Pred. No. 0.072;		
Matches	8;	Conservative	0;	Mismatches 3;
				Indels 0;
				Caps 0;
OY	1 SRXHXSMEXR 11			
DB	23 SRXHXSMEXR 33			
RESULT 2				
ID	Q95367	PRELIMINARY;	PRT;	450 AA.
AC	Q95367			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			

```

DE CBF1 interacting corepressor CIR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9903483; PubMed=9874765;
RA Hsieh J.D., Zhou S., Chen L., Young D.B., Hayward S.D.;
RT "CIR, a corepressor linking the DNA binding factor CBF1 to the histone
deacetylase complex."
RL Proc. Natl. Acad. Sci. U.S.A. 96:23-28(1999).
DR EMBL: AF086297; APO5243.1;
SQ SEQUENCE 450 AA; 52328 MW; 96AABF69CB73357 CRC64;

Query Match 74.5%; Score 35; DB 4; Length 450;
Best Local Similarity 63.6%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 359 SRTXHSPEKR 369

RESULT 3
ID 012804 PRELIMINARY; PRT; 451 AA.
AC 012804;
DT 01-NOV-1996 (TRMBLrel. 01, Created)
DT 01-NOV-1996 (TRMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE Receptin.
GN RECEPTIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RA Chai K.X., Li L., Chao J., Chao L.;
RT "Receptin: a novel human liver cDNA encoding a serpin-like molecule."
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: U03644; AAA17853.1;
DR InterPro: IPR000345; Cytochrome_b1nd.
DR PROSITE: PS00190; CYTOCHROME_C: UNKNOWN_1.
SQ SEQUENCE 451 AA; 52942 MW; 157F04804DA0817E CRC64;

Query Match 74.5%; Score 35; DB 4; Length 451;
Best Local Similarity 63.6%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 360 SRTXHSPEKR 370

RESULT 4
ID 08XBC4 PRELIMINARY; PRT; 672 AA.
AC 08XBC4;
DT 01-MAR-2002 (TRMBLrel. 20, Created)
DT 01-MAR-2002 (TRMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE Hydrogenase 4 membrane subunit.
GN HYB OR 23742 OR EGS3344.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;

```

```

RX MEDLINE=21074935; PubMed=11206551;
RA Ferra N.T., Plunkett G. III, Burland V., Mau B., Giasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posifai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIND 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL: AB005478; AAG57592.1;
DR EMBL: AP002561; BAB36767.1;
DR InterPro: IPR003918; NADph_oxred4.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; Oxidored_q1.1.
DR PRINTS: PR01437; NOXDRDPA54.
SQ SEQUENCE 672 AA; 72659 MW; D983361A569AC70E CRC64;

Query Match 72.3%; Score 34; DB 16; Length 672;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 9
Db 364 SRTXHSPEKR 372

RESULT 5
ID 09RH19 PRELIMINARY; PRT; 340 AA.
AC 09RH19;
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE Cytochrome oxidase D subunit B.
GN CYDB.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
Zymomonas.
NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ZM4;
RA Lee J.S., Kang H.S.;
RT "Sequence analysis of 6563 cosmid clone of Zymomonas mobilis ZM4
containing rna operon."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF088897; AAF16273.1;
DR InterPro: IPR003317; Cyto_ox_2.
DR Pfam: PF02322; Cyto_ox_2.1.
DR TIGRfams: TIGR00203; CYDB; 1.
SQ SEQUENCE 340 AA; 37206 MW; 3D7038A3F72EC316 CRC64;

Query Match 70.2%; Score 33; DB 2; Length 340;
Best Local Similarity 60.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 SRXHXSMEXR 11
Db 107 RVAAHSKEAR 116

```

RESULT 6

Q9DE12 PRELIMINARY; PRT; 4133 AA.
 AC Q9DE12; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DN DNA-dependent protein kinase catalytic subunit.
 GN DNA-PKCS.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LYMPHOCYTE;
 RX MEDLINE=20456723; PubMed=11003390;
 RA Fujimori A., Tatsumi K., Fukumura R., Ohnata T., Takahashi H.,
 Kawahara A., Tatsumi K., Abe M.;
 RT Identification of four highly conserved regions in DNA-PKCS.;
 RL Immunogenetics 51:965-973(2000).
 DR EMBL; AB016240; BAA36956.1;
 DR InterPro; IPR003151; FAT.
 DR InterPro; IPR003152; FATC.
 DR InterPro; IPR004003; PI3_P14_Kinase.
 DR Pfam; PF02259; FAT; 1.
 DR Pfam; PF02260; FATC; 1.
 DR Pfam; PF00454; PI3_P14_Kinase; 1.
 DR SMART; SM00146; PI3Kc; 1.
 DR PROSITE; PS00915; PI3_4_KINASE_1; UNKNOWN_1.
 DR PROSITE; PS02890; PI3_4_KINASE_3; 1.
 KW kinase.
 SQ SEQUENCE 4133 AA; 472522 MW; 0A156C131466A068 CRC64;
 Query Match 70.2%; Score 33; DB 13; Length 4133;
 Best Local Similarity 75.0%; Pred. No. 2; Be+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RXHXHME 9
 DB 1766 RDRHSMX 1773
 RESULT 7
 O8OGX4 PRELIMINARY; PRT; 4134 AA.
 AC O8OGX4; 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DN DNA-dependent protein kinase catalytic subunit.
 GN PRKDC.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21829525; PubMed=11839092;
 RX Fujimori A., Hashimoto H., Araki R., Saito T., Sato S., Kasama Y.,
 Tatsumi Y., Mori M., Fukumura R., Ohnata T., Tatsumi K., Abe M.;
 RT Sequence Analysis of 193.4 and 83.9 kbp of Mouse and Chicken Genomic
 DNAs Containing the Entire Prkdc (DNA-PKcs) Gene.;
 RL Radiat. Res. 157:298-305(2002).
 DR EMBL; AB028136; BAB91148.1;
 KW kinase.
 SQ SEQUENCE 4134 AA; 472647 MW; 8907B1E6E7E9A8 CRC64;
 Query Match 70.2%; Score 33; DB 13; Length 4134;
 Best Local Similarity 75.0%; Pred. No. 2; Be+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RXHXHME 9
 DB 1767 RDRHSMX 1774

RESULT 8

O9HNP4 PRELIMINARY; PRT; 196 AA.
 AC O9HNP4; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DN Vng2006C.
 GN VNG2006C.
 OS Halobacterium sp. (strain NRC-1).
 CC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Halobacterium.
 NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20504483; PubMed=11016950;
 RX Ng W.V., Kennedy S.P., Mahitras G.G., Bergquist B., Pan M.,
 Shukla H.D., Laskey S.R., Baliga N.S., Thorsson V., Strogina J.,
 Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
 Lelthausen B., Keller K., Cruz R., Denson M.J., Hough D.W.,
 Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 Alan M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 Ebdardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AF005094; AAC20176.1;
 KW Complete proteome.
 SQ SEQUENCE 196 AA; 21450 MW; 2F97F424E87C9E73 CRC64;

Query Match 68.1%; Score 32; DB 17; Length 196;
 Best Local Similarity 54.5%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXHMEXR 11
 DB 112 SRXHXHMEXR 122

RESULT 9

O9DBA2 PRELIMINARY; PRT; 277 AA.
 AC O9DBA2; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DN 1700023B02RLK protein.
 GN 1700023B02RLK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
 RX MEDLINE=21085660; PubMed=11217851;
 Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kato K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schirral L.M., Staib J.F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustinlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee R.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK008242; BAB25553.1; -
 DR MGD; MGI:1914185; 1700023B02R1K.
 DR PRINTS; PRO1574; TUBBYPROTEIN.
 SQ SEQUENCE 277 AA; 31797 MW; 4BE4D95516FBAC32 CRC64;

Query Match 68.1%; Score 32; DB 11; Length 277;
 Best Local Similarity 54.5%; Pred. No. 31;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SRXHSMEKX 11
 Db 200 SRSHRSPK 210

RESULT 10
 Q9DAE5 PRELIMINARY: PRT: 286 AA.
 AC Q9DAE5;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 1700023B02R1K protein.
 GN 1700023B02R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Iizawa M., Nishi K., Kiyosawa B., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischman W., Gaasterland T., Glasl C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK005900; BAB24307.1; -
 DR MGD; MGI:1914185; 1700023B02R1K.
 DR PRINTS; PRO1574; TUBBYPROTEIN.
 SQ SEQUENCE 286 AA; 32818 MW; 068003C5E894827B CRC64;

Query Match 68.1%; Score 32; DB 11; Length 286;
 Best Local Similarity 54.5%; Pred. No. 32;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 SRXHSMEKX 11
 Db 209 SRSHRSPK 219

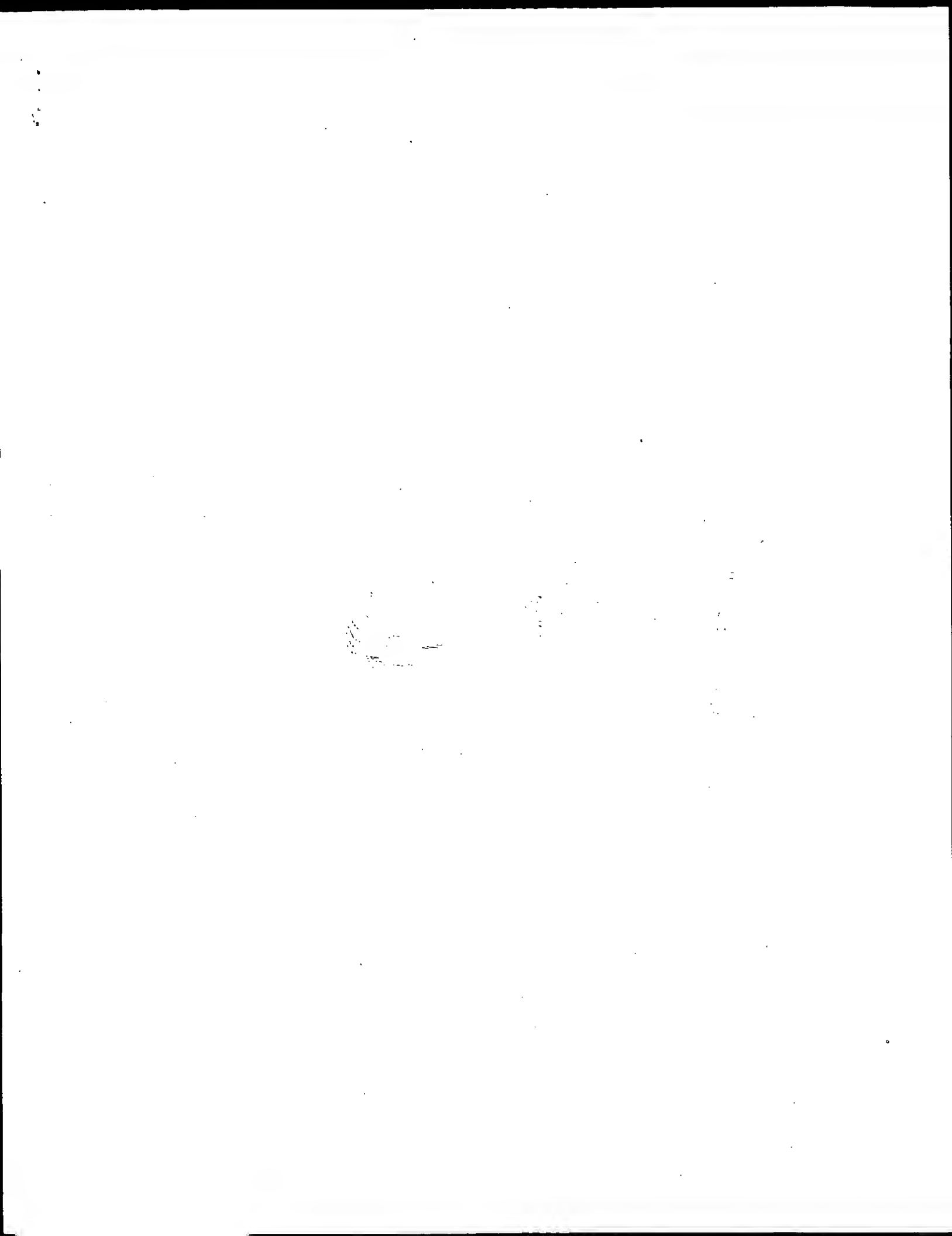
RESULT 11
 Q9DA19 PRELIMINARY: PRT: 383 AA.
 AC Q9DA19;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Phospholipase D endonuclease superfamily.
 GN Cr154.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/CX;
 RX MEDLINE=99000609; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Naratue R., Aravind L.,
 RA Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.,
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 DR EMBL: AF001289; AAC67745.1; -
 DR InterPro; IPR001736; PLD.
 DR Pfam; PF00614; PLDC; 2.
 DR SMART; SM00155; PLDC; 2.
 RN Endonuclease; Complete proteome.
 SQ SEQUENCE 383 AA; 43383 MW; 48F73A163553D2FA CRC64;

Query Match 68.1%; Score 32; DB 16; Length 383;
 Best Local Similarity 54.5%; Pred. No. 43;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 SRXHSMEKX 11
 Db 18 SRSHRSPK 28

RESULT 12
 Q9DA19 PRELIMINARY: PRT: 450 AA.
 AC Q9DA19;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 1700023B02R1K protein.
 GN 1700023B02R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Iizawa M., Nishi K., Kiyosawa B., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischman W., Gaasterland T., Glasl C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA Hayashizaki Y.;
 RP "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK006260; BAB2448.1;
 DR MGD; MGI:1914185; 1700023B02R1.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
 SO SEQUENCE 450 AA; 51852 MW; F32FL1BR6D6A4EAC CRC64;
 Query Match 68.1%; Score 32; DB 11; Length 450;
 Best Local Similarity 54.5%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 OY 1 SRXRXSHKEXR 11
 DB 373 SRSHSHSPK 383
 RESULT 13
 ID O9X0M5 PRELIMINARY; PRT; 648 AA.
 AC O9X0M5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE AMP-dependent DNA helicase.
 GN TM1238.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB / DSM 3109;
 RX MEDLINE=9287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Mark J.A., Liner K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1999).
 DR EMBL; AF001779; AAD3613.1;
 DR HSP; P56255; IPR.
 DR TIGR; TM1238;
 DR InterPro; IPR000212; UVD-helicase.
 DR Pfam; PF00580; UVD-helicase; 1.
 SO SEQUENCE 648 AA; 75842 MW; 427AFFE319250280 CRC64;
 Query Match 68.1%; Score 32; DB 16; Length 648;
 Best Local Similarity 62.5%; Pred. No. 72;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 2 RXRXHSMKE 9
 DB 370 RSHSHSL 377
 RESULT 14
 ID O8Y2S0 PRELIMINARY; PRT; 712 AA.
 AC O8Y2S0;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Probable acyl-CoA carboxylase alpha chain protein.
 GN RSC0265 OR RS05771.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salenouat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlet M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandelier M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Laye M., Moisan A., Robert C., Saurin M., Schlex T.,
 RA Sigler P., Theault P., Whalen M., Winkler P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RL Nature 415:497-502(2002).
 DR EMBL; AL646058; CAD13793.1;
 DR InterPro; IPR000089; Biotin_11poy1.
 DR InterPro; IPR000901; CPSase.
 DR Pfam; PF02785; biotin_carb_C; 1.
 DR Pfam; PF0289; CPSase_L_D2; 1.
 DR Pfam; PF02786; CPSase_L_D2; 1.
 DR PROSITE; PS00866; CPSASE_1; 1.
 DR PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
 KW Complete proteome.
 SO SEQUENCE 712 AA; 76240 MW; 37D79D7B8C222797 CRC64;
 Query Match 68.1%; Score 32; DB 16; Length 712;
 Best Local Similarity 50.0%; Pred. No. 80;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 OY 2 RXRXHSMKE 11
 DB 366 RIRGALR 375
 RESULT 15
 ID O8WM01 PRELIMINARY; PRT; 141 AA.
 AC O8WM01;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Transformer-2 alpha (Fragment).
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HYPOPHALMOS;
 RA Mungenast A.E., Ojeda S.R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY072878; AAL68856.1;
 FT NON_TER 1
 FT NON_TER 1
 SO SEQUENCE 141 AA; 15974 MW; AAB9A50C50670CA5 CRC64;
 Query Match 66.0%; Score 31; DB 6; Length 141;
 Best Local Similarity 54.5%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 SRXRXSHKEXR 11
 DB 50 SRSHSHSHRR 60
 Search completed: May 1, 2003, 14:33:01
 Job time : 16.0312 secs



GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 3.60938 Seconds

(without alignments)
126.404 Million cell updates/sec

Title: US-09-446-543a-74

Perfect score: 47

Sequence: 1 SRXHXHSEXR 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	91.5	83	1	PRRP_RAT
2	42	89.4	87	1	PRRP_HUMAN
3	42	89.4	98	1	PRRP_BOVIN
4	34	72.3	672	1	HYFB_ECOLI
5	33	70.2	497	1	YV80_MYCTU
6	32	68.1	452	1	YV84_MYCTU
7	32	68.1	460	1	VP41_BPAPS
8	31	66.0	277	1	MCRA_ECOLI
9	31	66.0	282	1	TRZA_HUMAN
10	31	66.0	495	1	HSP1_ARATH
11	30	63.8	333	1	ILVC_MYCAV
12	30	63.8	333	1	ILVC_MYCTU
13	30	63.8	340	1	ILVC_CAME
14	30	63.8	772	1	SM3A_RAT
15	30	63.8	1224	1	MSW5_YEAST
16	29	61.7	145	1	OAL_ONCOV
17	29	61.7	153	1	IEK1_MOUSE
18	29	61.7	219	1	LOT2_MOUSE
19	29	61.7	405	1	ITSP_MOUSE
20	29	61.7	462	1	US29_HCMYA
21	29	61.7	513	1	REP_HUMAN
22	29	61.7	522	1	REP_MOUSE
23	29	61.7	774	1	LOL2_HUMAN
24	29	61.7	915	1	Y007_CAEBL
25	29	61.7	2869	1	RBP1_PLAIV
26	28	59.6	277	1	T2PB_DROSE
27	28	59.6	319	1	TRXB_BUCAP
28	28	59.6	331	1	UL38_HCMYA
29	28	59.6	338	1	ILVC_CORGL
30	28	59.6	340	1	ILVC_STERR
31	28	59.6	363	1	ODPA_SMITA
32	28	59.6	388	1	ODPT_HUMAN
33	28	59.6	389	1	ODPA_ARATH

34	28	59.6	389	1	ODPA_PIG	P29804 sus acrofa
35	28	59.6	390	1	ODPA_HUMAN	P08559 homo sapien
36	28	59.6	390	1	ODPA_MOUSE	P35486 mus musculu
37	28	59.6	390	1	ODPA_RAT	P26288 rattus norv
38	28	59.6	391	1	ODPA_SOLIU	P32903 solianum tub
39	28	59.6	391	1	ODPT_MOUSE	P35487 mus musculu
40	28	59.6	391	1	ODPT_RAT	P06437 rattus norv
41	28	59.6	397	1	ODPA_CAEBL	P52899 caenorhabdl
42	28	59.6	397	1	ODPA_PEA	P52902 plium sativ
43	28	59.6	507	1	C7C3_ARATH	064638 arabidopsis
44	28	59.6	521	1	IMDH_CHYVI	050316 chlorobium
45	28	59.6	576	1	SYR_YERPE	Q8277 yersinia pe

ALIGNMENTS

RESULT 1
PRRP_RAT
ID PRRP_RAT STANDARD; PRT: 83 AA.
AC P81278:
DI 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRRP1; Proactin-releasing peptide PRRP2].
DE PRR.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. NCBI_Taxid=10116;
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE-98268781; PubMed-9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kikada C., Masuo Y., Amano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H.,
RA "A proactin-releasing peptide in the brain";
RA Nature 393:272-276(1998).
RT [2]
RN TISSUE SPECIFICITY.
RX MEDLINE-9942652; PubMed-10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H.,
RA Sunino Y., Fujino M.,
RA "Tissue distribution of proactin-releasing peptide (PRP) and its receptor";
RL Regul. Pept. 83:1-10(1999).
CC -I- FUNCTION: Stimulates proactin (PRP) release and regulates the expression of proactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -I- TISSUE SPECIFICITY: widely expressed, with highest levels in medulla oblongata and hypothalamus.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/or-send-an-email-to-license@isb-sib.ch>).
CC EMBL: AB015418; BAA29026.1;
DR Hormone; Amidation; Signal; Cleavage on pair of basic residues.
KW SIGNAL.
FT 1
FT SIGNAL.
FT 21
FT PEPTIDE 22
FT 52
FT PEPTIDE 33
FT 52
FT MOD_RES 52
FT 52
SQ SEQUENCE 83 AA; 9215 MW; DOC75A264EEAF29 CRC64;
Query Match 91.5%; Score 43; DB 1; Length 83;

Best Local Similarity 72.7%; Pred. No. 0.0078;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
DB 22 SRAHSHMEIR 32

RESULT 2

PRRP_HUMAN STANDARD; PRT; 87 AA.
AC P81277;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRRP31; Proactin-releasing peptide PRRP20].
GN PRH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98268781; Pubmed=9607765;
RA Hinuma S., Nabata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=99426652; Pubmed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H.,
RT "Tissue distribution of prolactin-releasing peptide (PrRP) and its receptor.";
RL Regul. Pept. 83:1-10(1999).
CC -1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: AB015419; BAA29027.1;
DR KIM; 602663;
KW Hormone; Amidation; Signal.
FT SIGNAL; 1
FT PEPTIDE 23 53 BY SIMILARITY.
FT PEPTIDE 34 53 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT MOD_RES 53 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT SEQUENCE 87 AA; 9639 MW; 229A2F3F50CF981B CRC64; AMIDATION (G-54 PROVIDE AMIDE GROUP).

Query Match 89.4%; Score 42; DB 1; Length 87;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
DB 23 SRAHSHMEIR 33

RESULT 3

PRRP_BOVIN STANDARD; PRT; 98 AA.
AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRRP31; Proactin-releasing peptide PRRP20].
GN PRH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC TISSUE=Brain;
RX MEDLINE=98268781; Pubmed=9607765;
RA Hinuma S., Nabata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=99426652; Pubmed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H.,
RT "Tissue distribution of prolactin-releasing peptide (PrRP) and its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: AB015417; BAA29025.1;
DR KIM; 602663;
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL; 1
FT PEPTIDE 23 53 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 53 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
FT SEQUENCE 98 AA; 10544 MW; 08AC35A13B0F908 CRC64;

Query Match 89.4%; Score 42; DB 1; Length 98;
Best Local Similarity 72.7%; Pred. No. 0.015;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
DB 23 SRAHSHMEIR 33

RESULT 4

HYRB_ECOLI STANDARD; PRT; 672 AA.
AC P23482;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase-4 component B (Ec 1.1.1.1).
GN HYRB OR B2482.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=K12;
RA Andrews S.C., McClellan J., Ambler A., Quail M., Berks B.C., Guest J.R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

RESULT 4

RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Colado-Vides J., Glander J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Man B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97349980; PubMed-9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunishi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Ohshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,
 RA Yamagata S., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [4]
 RP SEQUENCE OF 1-34 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-91202105; PubMed-2016588;
 RA Andrews S.C., Harrison P.M., Guest J.R.;
 RT "A molecular analysis of the 53.3 minute region of the *Escherichia*
 RT coli linkage map.";
 RL J. Gen. Microbiol. 137:361-367(1991).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (by similarity).
 CC -1- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
 CC OF CHLOROPLASTS OR MITOCHONDRIA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M63654; AAB8564.1;
 DR EMBL: AE000335; AAC7535.1;
 DR EMBL: D90876; BAA16360.1;
 DR EMBL: D90877; BAA16370.1;
 DR PIR: D49749; D49749.
 DR Ecogene: EGI1282; hyfB.
 DR InterPro: IPR001750; Oxidored_q1.
 DR Pfam: PF00361; oxidored_q1; 1.
 KW Oxidoreductase; Transmembrane; Inner membrane; Complete proteome.
 DR TRANSMEM 6 26
 FT TRANSMEM 31 51 POTENTIAL.
 FT TRANSMEM 80 100 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 165 185 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 428 448 POTENTIAL.
 FT TRANSMEM 475 495 POTENTIAL.
 FT TRANSMEM 505 525 POTENTIAL.
 FT TRANSMEM 532 552 POTENTIAL.
 FT TRANSMEM 652 672 POTENTIAL.
 SQ SEQUENCE 672 AA; 72582 MW; 3CABC2BD25E54B1B CRC64;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 9
 DB 364 SRXHXSMEXR 372
 RESULT 5
 ID Y880.MVCTU STANDARD; PRT; 497 AA.
 AC 006343;
 DT 30-MAY-2000 (rel. 39, Created)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Hypothetical protein RV3480c.
 GN RV3480C OR MT3584 OR MCVY13B12.33C.
 OS *Mycobacterium tuberculosis*.
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; *Mycobacterium*.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37Rv;
 RX MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jorgels K., Krogh A., McLaren J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail A.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Esmailova M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J.A., Khouri E., Gill J., Mikula A.,
 RA Bishop W.;
 RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
 RT laboratory strains.";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0089 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z95390; CAB08703.1;
 DR EMBL: AE007161; AAK47943.1; ALT_INIT.
 DR TIGR: MT3584;
 DR Tuberculist: RV3480c;
 DR InterPro: IPR004255; UPF0089.
 DR Pfam: PF03007; UPF0089; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 427 427 L -> V (IN REF. 2).
 SQ SEQUENCE 497 AA; 53314 MW; 90D661E32B5558C2 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 497;
 Best Local Similarity 45.5%; Pred. No. 8;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 DB 108 SRXHXSMEXR 118

RESULT 6
YH4_YEAST STANDARD; PRT; 452 AA.
ID YH4_YEAST
AC P32047;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 49.2 kDa protein IMP4-MSK1 intergenic region.
GN YH4074C OR N3359 OR YH41.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96267764; PubMed=8701611;
RA Poehlmann R., Philippson P.;
RT *Sequencing a cosmid clone of Saccharomyces cerevisiae chromosome XIV
RT reveals 12 new open reading frames (ORFs) and an ancient duplication
RT of six ORFs.*;
RL Yeast 12:391-402(1996).
RN [2]
RP SEQUENCE OF 1-217 FROM N.A.
RX MEDLINE=91202535; PubMed=2016746;
RA Gatti D., Tzagoloff A.;
RT *Structure and evolution of a group of related aminoacyl-tRNA
RT synthetases.*;
RL J. Mol. Biol. 218:557-568(1991).
CC -1 SIMILARITY: TO YEAST YIL135C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X86470; CA60186.1; -;
DR EMBL; Z71350; CA93848.1; -;
DR EMBL; X57360; -; NOT_ANNOTATED_CDS.
DR PIR; S27290; S27290.
DR SGD; S0005018; YNL074C.
KW Hypothetical protein.
FT DOMAIN 211 214 POLY-THR.
FT DOMAIN 271 274 POLY-SER.
SQ SEQUENCE 452 AA; 49210 MW; 9C4FF02D38D6963B CRC64;
Query Match 68.1%; Score 32; DB 1; Length 452;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SRKXHSMEXR 11
DB 189 SRSHSRAATR 199

RESULT 7
VP41_BPAPS STANDARD; PRT; 460 AA.
ID VP41_BPAPS
AC Q9T107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative protein P41.
GN 41.
OS Bacteriophage ABSE-1.
OC Viruses; dsDNA viruses, no RNA stage; Caulovirales; Podoviridae;
CC Unclassified Podoviridae.
OX NCBI_TaxID=106199;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99420383; PubMed=10489345;
RX van der Wilk F., Dullemeers A.M., Verbeek M., van den Heuvel J.F.J.M.;
RT "Isolation and characterization of ABSE-1, a bacteriophage infecting
RT the secondary endosymbiont of acyrthosiphon pisum.*";
RL Virology 262:104-113(1999).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF157835; AAF03984.1; -;
KW Hypothetical protein.
SQ SEQUENCE 460 AA; 51920 MW; 9572842E1FFDE31 CRC64;
Query Match 68.1%; Score 32; DB 1; Length 460;
Best Local Similarity 55.6%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 SRKXHSME 9
DB 259 ARAHYNAME 267

RESULT 8
MCRA_ECOLI STANDARD; PRT; 277 AA.
ID MCRA_ECOLI
AC P24200;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methylcytosine-specific restriction enzyme A (EC 3.1.21.-)
DE (EcoMCRa).
GN MCRA OR RGIA OR B1159.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041637; PubMed=1938927;
RA Hion K.J., Sedgwick S.G.;
RT *Cloning and structural characterization of the mcra locus of
RT Escherichia coli.*;
RL J. Bacteriol. 173:7368-7373(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Ramalingam R., Prasad R., Shivapriya R., Dharmalingam K.;
RT *Molecular cloning and sequencing of mcra locus and identification of
RT Mcra protein in Escherichia coli.*;
RL J. Biosci. 17:217-232(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick N.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.*";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alts H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kawai K., Kasahimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.,
 RT "A 718-bp DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-156(1996).
 CC -1- FUNCTION: RESTRICTION OF 5-METHYL AND 5-HYDROXYMETHYLCYTOSINES AT
 CC THE SPECIFIC DNA SEQUENCE C(M)CGG.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: W76667; AAA68481.1;
 DR EMBL: Z19104; CAA79520.1;
 DR EMBL: A800215; AAC74243.1;
 DR EMBL: D90750; BAA35995.1;
 DR PIR: A41424; A41424.
 DR REBASE: 2832; ECKRCRA.
 DR Ecogene: EGI0573; MCR.
 DR InterPro: IPR002711; HNH.
 DR InterPro: IPR003615; HNH_nuc.
 DR Pfam: PF01844; HNH_1.
 DR SMART: SM00507; HNHc_1.
 DR Hydroxylase; Endonuclease; Nuclease; Restriction system;
 KW Complete proteome.
 FT COMPLETE 214
 FT SEQUENCE 277 AA; 31390 MW; E5F2627DFDEC402 CRC64;
 SQ
 Query Match 66.0%; Score 31; DB 1; Length 277;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 RXHXSHMR 11
 DB 69 RKHMSLDR 78
 RESULT 9
 TR2A_HUMAN STANDARD; PRT; 282 AA.
 ID Q13595;
 AC Q13595;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transformer-2 protein homolog (TRA-2 alpha).
 GN TRA2A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96392356; PubMed=8799144;
 RA Davelider B., Amaya-Manzanares F., Mattox W.;
 RT "A human homologue of the *Drosophila* sex determination factor
 RT transformer-2 has conserved splicing regulatory functions.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:9004-9009(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=uterus;
 RA Strausberg R.;
 RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION, SUBCELLULAR LOCATION, AND PHOSPHORYLATION.
 RX TISSUE=Cervical carcinoma;
 RA MEDLINE=98206475; PubMed=9546399;
 RT Tacke R., Tohyama M., Ogawa S., Manley J.L.;
 RT "Human Tra2 proteins are sequence-specific activators of pre-mRNA

RT splicing.";
 RL Cell 93:139-148(1998).
 CC -1- FUNCTION: Sequence-specific RNA-binding protein which participates
 CC in the control of pre-mRNA splicing.
 CC -1- SUBUNIT: Binds to A3 enhancer proteins Srp75, Srp55, Srp40 and
 CC Srp30.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- PTM: Phosphorylated in the RS domains.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U53209; AAC50658.1;
 DR EMBL: BC017094; AAH17094.1;
 DR HSRP; P19339; 2SXL.
 DR MIM: 602718;
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM03102; RRM; 1.
 DR PROSITE: PS00303; RRM_NP_1;
 DR PROSITE: PS00303; RRM_NP_1;
 DR RNA-binding; mRNA splicing; mRNA processing; Alternative splicing;
 KW Nuclear protein; Phosphorylation.
 FT DOMAIN 30 112 ARG/SER-RICH (RS1 DOMAIN).
 FT DOMAIN 119 197 RNA-BINDING (RRM).
 FT DOMAIN 198 225 LINKER.
 FT DOMAIN 226 282 ARG/SER-RICH (RS2 DOMAIN).
 FT VARSPLIC 1 101 MISSING (IN SHORT ISOFORM).
 FT VARSPLIC 214 214 H -> Q (IN SHORT ISOFORM).
 FT VARSPLIC 215 282 MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 282 AA; 32688 MW; EDB5ABE7BEA023FD CRC64;
 SQ
 Query Match 66.0%; Score 31; DB 1; Length 282;
 Best Local Similarity 54.5%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SRXSHMR 11
 DB 71 SRSRSHMR 81
 RESULT 10
 HSFL_ARATH STANDARD; PRT; 495 AA.
 ID P4151; Q23615;
 AC P4151; Q23615;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Heat shock factor protein 1 (HSF 1) (Heat shock transcription factor
 DE 1) (HSF 1).
 GN HSF1 OR A4617750 OR D44910C.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95036006; PubMed=7948881;
 RA Huebel A., Schoeffl F.;
 RT "Arabidopsis heat shock factor: isolation and characterization of the
 RT gene and the recombinant protein.";
 RL Plant Mol. Biol. 26:353-362(1994).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV, Columbia;
 RX MEDLINE=98121113; PubMed=9461215;
 RA Bevan M., Bancroft I., Bent E., Loye K., Goodman H., Dean C.,
 RA Bergkamp R., Dirksen W., van Staveren M., Stiekema W., Drost L.,
 RA Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,
 RA Wedler B., Wedler E., Wambutt R., Weitzenecker T., Pohl T., Terry N.,
 RA Gleier J., Villarroel R., de Clerck R., van Montagu M., Lecharny A.,
 RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
 RA Ertan K.-D., Rieger M., Scheffer M., Funk B., Mueller-Auer S.,
 RA Siveley M., James R., Montfort A., Pons A., Pulgomech P., Douka A.,
 RA Vokelaitis E., Milioni D., Natopoulos P., Piravandi E., Obermaier B.,
 RA Hilbert H., Duesterhoef A., Moores T., Jones J.D.G., Eneva T.,
 RA Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
 RA Delseny M., Voet M., Volckaert G., Mewes H.-W., Klosterman S.,
 RA Schueller C., Chalwatzis N.;
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 RT Arabidopsis thaliana.";
 RL Nature 391:485-488(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mohl K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoef A., Stiekema W., Ertan K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Griwell L.A., Rieger H.,
 RA Wachelgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delseny M., Pulgomech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnel J., Zimmermann W., Wedler B., Ridley P.,
 RA Langham S.-A., McCullagh B., Bihem L., Robben J.,
 RA van der Schueren J., Grymonprez B., Chuang Y.-J., Vandebussche F.,
 RA Breken M., Weltens J., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Nibbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirksen W.,
 RA Moeljan P., Klein Lankhorst R., Rose M., Half J., Koetter P.,
 RA Benneiser S., Nempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA de Keyser A., Buyschaert C., Gielen J., Villarroel R., de Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Nall S., Kay M., Leonard N., McIay K., Mayes R.,
 RA Pettett A., Ratendream M.A., Lyne M., Benes V., Rechman S.,
 RA Borkova D., Blocher H., Scharfe M., Grimm M., Lochert T.-H.,
 RA Dose S., de Haan M., Hartse A.C., Schaefer M., Kellner-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandtath K., Dauner D., Heitzl A.,
 RA Neumann S., Argilou A., Vitale D., Liuzzi R., Piravandi E.,
 RA Massenot O., Quigley F., Clapaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chefor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandembol H., Barges H., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heljnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dechta N., Gnoj L., Schutz K., Huang E., Spiegler L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stonewall T., Rakici J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton J., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spleth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Dione K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun N., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Nagesawa A., Hameed A., Iodhi M., Johnson A.,
 RA Chen E., Barra M., Martensen R., McComble W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
 CC PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER.

CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: EXHIBITS TEMPERATURE-DEPENDENT PHOSPHORYLATION
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HSF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC
 CC EMBL: X76167; CAA53761.1; -
 CC EMBL: Z97344; CAB10555.1; -
 CC EMBL: AL161547; CAB78778.1; -
 CC PIR: S38873; S38873.
 CC NSSP: P22121; HSPF.
 CC TRANSFAC: T04394; -
 CC InterPro: IPR000232; NSF_DNA_Bind.
 CC InterPro: IPR002341; HSF_FTS.
 CC Pfam: PF00447; HSF_DNA_Bind. 1.
 CC PRINTS: PR00056; NSPDOMAIN.
 CC PRODOM: PD001788; HSF_DNA_Bind; 1.
 CC SMART: SM00415; HSF_1.
 CC PROSITE: PS00434; HSF_DOMAIN; 1.
 CC Transcription regulation; Nuclear protein; DNA-binding; Activator;
 CC Phosphorylation; Heat shock; Multigene family.
 CC
 CC FT DNA_BIND 50 144
 CC FT DOMAIN 150 153
 CC FT POLY-SER.
 CC FT POLY-GLN.
 CC FT DOMAIN 204 207
 CC FT POLY-ASN.
 CC FT CONFLICT 216 216 L -> M (IN REF. 1).
 CC FT CONFLICT 285 285 E -> D (IN REF. 1).
 CC FT CONFLICT 336 336 I -> T (IN REF. 1).
 CC FT CONFLICT 390 390 F -> Y (IN REF. 1).
 CC FT CONFLICT 462 473 MISSING (IN REF. 1).
 CC FT SEQUENCE 495 AA; 55744 MW; 2F0DE5252B5682A CRC64;
 CC
 CC Query Hatch 66.0%; Score 31; DB 1; Length 495;
 CC Best local similarity 55.6%; Pred. No. 21;
 CC Matches 5; Conservative 1; Mismatches 3; Gaps 0;
 CC
 CC QY 1 SRXKXNSME 9
 CC DB 277 SNSSHSL 285
 CC
 CC RESULT 11
 CC ILVC_MYCAN STANDARD; PRT; 333 AA.
 CC AC 059500;
 CC DT 01-NOV-1997 (rel. 35, Created)
 CC DT 01-NOV-1997 (rel. 35, Last sequence update)
 CC DT 15-JUN-2002 (rel. 41, Last annotation update)
 CC DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Aceto-hydroxy-acid
 CC isomerase) (Alpha-keto-beta-hydroxyacid reductoisomerase).
 CC GN ILVC.
 CC OS Mycobacterium avium.
 CC CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1764;
 CC
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=97080504; PubMed=8921849;
 CC RA Gusberti L., Cantoni R., de Rossi E., Branzoni M., Riccardi G.;
 CC RT "Cloning and sequencing of the ilvBNC gene cluster from Mycobacterium
 CC avium.";
 CC RL Gene 177:83-85(1996).
 CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
 CC - (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
 CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
 CC -1- SIMILARITY: BELONGS TO THE KETOL-ACID REDUCTOISOMERASE FAMILY.

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL021287; CAA16086.1; -
DR EMBL: AE007128; AAA47410.1; ALT_INIT.
DR HSSP: 001292; 10MG.
DR TIGR: MT3081; -
DR Tuberculist: RV3001c; -
DR InterPro: IPR000506; ACh_isomctcse.
DR Pfam: PF01450; llyvc; 1.
DR TIGRFAMs: TIGR00465; llyvc; 1.
KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
KW Complete proteome.
FT ACT_SITE 105 105 POTENTIAL.
SQ SEQUENCE 333 AA; 36090 MW; 49669E3EDACE25EC CRC64;

Query Match 63.8%; Score 30; DB 1; Length 333;
Best Local Similarity 45.5%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRKXKSMEXR 11
   1: | | | | |
Db 24 SQGHASHLSLR 34

RESULT 13
ILVC_CAMJF STANDARD; PRT; 340 AA.
ID ILVC_CAMJF
AC 09PHN5;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
DE isomeroreductase) (Alpha-Keto-beta-hydroxyacil reductoisomerase).
GN ILVC OR C30632.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OK NCBI_TaxID=157;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mangall K., Kelsey J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jagers K., Kariyeh A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -1- SIMILARITY: BELONGS TO THE KETO-ACID REDUCTOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL139075; CAB75268.1; -
DR InterPro: IPR000506; ACh_isomrctcse.
DR Pfam: PF01450; llyvc; 1.
DR TIGRFAMs: TIGR00465; llyvc; 1.
KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP.

```

KW Complete proteome. 109 POTENTIAL.
 FT ACT_SITE 109
 SQ SEQUENCE 340 AA; 36934 MW; DA556ED90C86C780 CRC64;
 Query Match 63.8%; Score 30; DB 1; Length 340;
 Best Local Similarity 45.5%; Pred. No. 23;
 Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 OY 1 SRXHSMEXR 11
 Db 28 SGRHMANLR 38

RESULT 14
 SM3A_RAT STANDARD; PRT; 772 AA.
 AC 063548;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DN Semaphorin 3A precursor (Semaphorin III) (Sema III).
 GN SEMA3A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=Master; TISSUE=Brain;
 RX MEDLINE=97073089; PubMed=8915837;
 RA Giger R.U., Wolter D.P., De Wit G.M.J., Verhaagen J.;
 RA "Anatomy of rat semaphorin III/collapsin-1 mRNA expression and
 RA relationship to developing nerve tracts during neuroembryogenesis.";
 RL J. Comp. Neurol. 375:318-392(1996).
 CC -1- FUNCTION: MAY BE INVOLVED IN GUIDING GROWING AXONS TOWARDS THEIR
 CC TARGETS BY FORMING A MOLECULAR BARRIER THAT INSTRUCTS AXONS TO
 CC ENGAGE IN THE FORMATION OF SPECIFIC NERVE TRACTS. BINDS TO
 CC NEUROFILIN.
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -1- DEVELOPMENTAL STAGE: AT E-11, EXPRESSION WAS RESTRICTED TO THE
 CC OFLACROFARY PIT, THE BASAL AND ROSTRAL SURFACE OF THE TELEENCEPHALIC
 CC VESICLE, THE EYE ANLAGE, THE EPITHELIUM OF RATNER'S POCH, AND
 CC SOMITES. AT LATER DEVELOPMENTAL STAGES, IT WAS WIDELY DISTRIBUTED
 CC IN NEURONAL AS WELL AS IN MESENCHYMAL AND EPITHELIAL STRUCTURES
 CC OUTSIDE THE NERVOUS SYSTEM. AFTER BIRTH, MESENCHYMAL LEVELS
 CC DECREASED RAPIDLY AND EXPRESSION BECAME RESTRICTED TO SPECIFIC
 CC SETS OF NEURONS IN THE CNS. IN THE MATURE CNS, IT IS DETECTABLE IN
 CC METABOLIC CELLS, NEURONS OF THE ACCESSORY BULB AND CEREBRAL CORTEX,
 CC CEREBELLAR PURKINJE CELLS, AS WELL AS A SUBSET OF CRANIAL AND
 CC SPINAL MOTONEURONS.
 CC -1- DOMAIN: STRONG BINDING TO NEUROFILIN IS MEDIATED BY THE CARBOXY
 CC THIRD OF THE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X95286; CAA64607.1; -
 DR InterPro: IPR003599; 1g.
 DR InterPro: IPR003006; 1g_MHC.
 DR InterPro: IPR003659; Pfam-11ke.
 DR Pfam: PF00047; 1g; 1.
 DR Pfam: PF01403; Sema; 1.
 DR SMART: SMO0409; IG; 1.
 DR SMART: SMO0423; PSI; 1.

KW Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
 KW Developmental protein; Glycoprotein.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 SEMAPHORIN 3A.
 FT DOMAIN 240 538 SEMA.
 FT DOMAIN 643 730 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 728 770 AMG/LYS-RICH (BASIC).
 FT DISULFID 650 723 BY SIMILARITY.
 FT CARBOHYD 53 53 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAc...) (POTENTIAL).
 SQ SEQUENCE 772 AA; 88808 MW; 240907812F9F2D2 CRC64;
 Query Match 63.8%; Score 30; DB 1; Length 772;
 Best Local Similarity 62.5%; Pred. No. 57;
 Matches. 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 4 HXHSMEXR 11
 Db 577 HCHSLER 584

RESULT 15
 MSNS_YEAST STANDARD; PRT; 1224 AA.
 ID MSNS_YEAST
 AC P52918;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MSNS protein.
 GN MSNS OR YDR335W OR D9651.5.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetiales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RA Alepuz P.M., Struch F.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Johnston K., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Kards E., Meneses S., Miller N., Nian M., Pauley A., Peluso D.,
 RA Wilcox L., Woldman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X93302; CAA63705.1; -
 DR EMBL: U51052; AAB64771.1; -
 DR SGD: S0002743; MSNS.
 FT CONFLICT 886 886 D -> G (IN REF. 1).
 FT CONFLICT 1011 1011 S -> R (IN REF. 1).
 SQ SEQUENCE 1224 AA; 142117 MW; C47A6767CBA44486 CRC64;
 Query Match 63.8%; Score 30; DB 1; Length 1224;
 Best Local Similarity 54.5%; Pred. No. 96;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 OY 1 SRXHSMEXR 11
 II : I I I I

Thu May 1 15:54:32 2003

Db 1014 SRLNNHOMEXR 1024

Search completed: May 1, 2003, 14:33:22
Job time : 5.60938 secs

us-09-446-543a-74.rsp

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 6.1875 Seconds

(without alignments)
170,906 Million cell updates/sec

Title: US-09-446-543a-74

Sequence: 1 SRXHXSMEXR 11

Scoring table: BLOSUM62
Gapop 10.0, Gapept 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	91.5	83	2	JC7607
2	35	74.5	451	2	G01227
3	34	72.3	672	2	A65024
4	34	72.3	672	2	H91046
5	34	72.3	672	2	D85891
6	33	70.2	497	2	E70568
7	32	68.1	196	2	D84351
8	32	68.1	383	2	F71550
9	32	68.1	452	2	S53906
10	31	66.0	448	2	G72279
11	31	66.0	277	2	A41424
12	31	66.0	495	1	S52641
13	30	63.8	330	2	H84075
14	30	63.8	332	1	B70015
15	30	63.8	333	1	UC5166
16	30	63.8	333	2	D70855
17	30	63.8	340	2	G81411
18	30	63.8	393	2	T28994
19	30	63.8	509	2	T36710
20	30	63.8	599	2	F82291
21	30	63.8	1224	2	S70100
22	29	61.7	122	2	G81198
23	29	61.7	122	2	G81775
24	29	61.7	153	2	S33363
25	29	61.7	194	2	A23277
26	29	61.7	292	2	T05317
27	29	61.7	307	2	T11104
28	29	61.7	462	1	Q08ED4
29	29	61.7	467	2	H88493

prolactin-releasing
receptor - human
Hydrogenase-4 comp
hydrogenase 4 memb
hypothetical prote
hypothetical prote
probable endonucle
YMK1 protein - yea
ATP-dependent DNA
5-methylcytosine-s
heat shock transcr
thioredoxin reduct
ketol-acid reduct
ketol-acid reduct
hypothetical prote
hypothetical prote
conserved hypothet
MSM protein - yea
hypothetical prote
gly96 protein - mo
gamma-secalin - ry
translational initia
HRRF4 protein - hu
Protein F57B9.10

30	29	61.7	506	2	S37583	RING finger protei
31	29	61.7	513	1	TVHDF	ret finger protein
32	29	61.7	569	2	F75381	probable two-compo
33	29	61.7	815	2	T15402	hypothetical prote
34	29	61.7	2829	2	A42771	reticulocyte-bind
35	28	59.6	23	2	A60226	pyruvate dehydroge
36	28	59.6	106	2	G87103	conserved hypothet
37	28	59.6	178	2	T49712	hypothetical prote
38	28	59.6	233	2	F87216	probable Ycrf-fam1
39	28	59.6	234	2	D70829	transcription init
40	28	59.6	277	2	S57336	regulatory protein
41	28	59.6	295	2	T48869	serotonin receptor
42	28	59.6	302	4	T02275	hypothetical prote
43	28	59.6	303	2	T47356	glycosyltransferas
44	28	59.6	303	2	A84114	probable N-acetyl
45	28	59.6	312	2	C71262	

ALIGNMENTS

RESULT 1

JC7607

prolactin-releasing peptide - rat

C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001

C:Accession: JC7607

Biochem. Biophys. Res. Commun. 281, 53-56, 2001

A:Title: Isolation and characterization of the rat prolactin-releasing peptide gene:

A:Reference number: JC7607; MUID:21092785; PMID:11178959

A:Contents: Spleen

A:Accession: JC7607

A:Molecule type: DNA

A:Residues: 1-83 <YAM>

A:Cross-references: DDBJ:AB040612; DDBJ:AB040613

C:Comment: This peptide induces arachidonic acid metabolite release from rat anterior

release, and stimulation of ACTH secretion from the pituitary.

C:Genetics:

A:Gene: PRRP

A:Introns: 33/1

Query Match

Best Local Similarity

Matches 8; Conservative

Score 43; DB 2; Length 83;

Pred. No. 0.037;

Mismatches 3; Indels 0; Gaps 0;

Query 1 SRXHXSMEXR 11

DB 22 SRXHXSMEXR 32

RESULT 2

G01227

receptor - human

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999

C:Accession: G01227

R:Chao, D.

submitted to the EMBL Data Library, November 1993

A:Reference number: G06292

A:Accession: G01227

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-451 <CHAS>

A:Cross-references: EMBL:U03644; NID:9476104; PIDD:AA17853.1; PTD:9476105

C:Genetics:

A:Gene: receptor

Query Match

Best Local Similarity

Matches 7; Conservative

Score 35; DB 2; Length 451;

Pred. No. 9.6;

Mismatches 4; Indels 0; Gaps 0;

Query 1 SRXHXSMEXR 11

A:Accession: J084351
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-196 <STOP>
 A:Cross-references: GB:AE004437; NID:g10581437; PIDN:AAG20176.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG2006C
 C:Superfamily: hypothetical protein MJ0690

Query Match	68.1%;	Score 32;	DB 2;	Length 196
Best Local Similarity	54.5%;	Pred. No. 17;		
Matches	6;	Conservative	1;	Mismatches 4;
				Indels

QY	1	SRHXHSMEXR	11
	11	1	1
Db	112	SRHAQSLDR	122

RESULT 8
 F71550
 Probable endonuclease Cti154 [similarity] - Chlamydia trachomatis (serotype D, strain UW-36)
 C.Species: Chlamydia trachomatis
 C.Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 04-Feb-2000
 C.Accession: F71550
 R.Stephenson, R.S.; Kalam, S.; Jammel, C.D.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, D.
 Science 282, 754-759, 1998
 A.Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A.Reference number: A15170; MUID:99000809; PMID:9744136
 A.Accession: F71550
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-383 <ARN>
 A.Cross-references: GB:AE001289; GB:AE001273; NID:g33328550; PIDN:AA67745.1; PID:g33328550
 A.Experimental source: serotype D, strain UW-3/Ck
 C.Genetics:
 A.Gene: Cti154

Query Match	68.1%;	Score 32;	DB 2;	Length 383;
Best Local Similarity	54.5%;	Pred. NO. 34;		
Matches	6;	Conservative	0;	Mismatches 5;
				Indels

```
QY      1 SRXHXHSMEXR 11
          | | | | |
Db      18 SRSHRHSSSR 28
```

RESULT 9
 S53906
 YKK1 protein - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein N2359; protein YNL074c
 C:Species: Saccharomyces cerevisiae
 C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
 C:Accession: S53906; S27290; S63006; S63935
 R:Poehlmann, R.; Philippsen, P.
 submitted to the EMBL Data Library, April 1995
 A:Reference number: S53896
 A:Accession: S53906
 A:Molecule type: DNA
 A:Residues: 1-452 <POE>
 A:Cross-references: EMBL:X66470; NID:g791101; PID:g791112
 R:Ritzgaard, A.
 submitted to the EMBL Data Library, January 1991
 A:Reference number: S27287
 A:Accession: S27290
 A:Molecule type: DNA
 A:Residues: 1-217 <TZA>
 A:Cross-references: EMBL:X57360
 R:Poehlmann, R.; Philippsen, P.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S62997
 A:Accession: S63006
 A:Molecule type: DNA
 A:Residues: 1-452 <POE>

A; Cross-references: EMBL:Z71350; NID:q1301960; PID:q1301961; MIPS:YML074C
A; Experimental source: strain S288C
R; Poehmann, R.; Philippson, P.

A;Gene: SGD:YMK1
A;Cross-references:
A;Map position: 14L

```
QY      1 SRHXNSMEXR 11
          ||| |
Db      189 SRSHSHSATR 199
```

RESULT 10
 G72279
 AMP-dependent DNA helicase - Thermotoga maritima (strain MSB)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: G72279
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hild
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference numbers: A12200; M01D:99287316; PMID:10360571
 A:Accession: G72279
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-648 <ARN>
 A:Cross-references: GB:AB001779; GB:AE000512; NID:g4981777; PIDN:AD36313.1; PID:g498
 A:Experimental source: strain MSB
 C:Genetics:
 A:Gene: TM1238
 C:Superfamily: helicase II

Query Match	68.1%;	Score 32;	DB 2;	Length 648;
Best Local Similarity	62.5%;	Pred. No. 58;		
Matches	5;	Conservative	1;	Mismatches 2;
				Indels

```
QY      2 RXHXHSM E 9
Db      370 RSHSHSL E 377
```

RESULT 11
 M1424
 5-methylcytosine-specific restriction enzyme A (EC 3.1.21.-) - *Escherichia coli* (strain
 C:Species: *Escherichia coli*
 C:Date: 21-Apr-1992 #sequence-revision 21-Apr-1992 #text-change 01-Mar-2002
 C:Accession: M1424; S48131; D64861
 R:Hlom, K.; Sedgwick, S.G.
 J. Bacteriol. 173, 7368-7373, 1991
 A>Title: Cloning and structural characterization of the *mcrA* locus of *Escherichia coli*
 A:Reference number: M1424; MUID:92041637; PMID:1938927
 A:Accession: M1424
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-277 <HIO>
 A:Cross-References: GB:M16667; NID:g146793; PIDN:AA68481.1; PID:g146794
 A:Experimental source: strain K-12

R.Ramalingam, R.; Prasad, R.; Shivapriya, R.; Dharmalingam, K.
 J. Biosci. 17, 217-232, 1992
 A>Title: Molecular cloning and sequencing of mcrA locus and identification of McrA prote
 A/Reference number: 548131
 A/Accession: 548131
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-277 <RAN>
 A/Cross-references: EMBL:219104; NID:941984; PIDD:CAV9520.1; PID:941985
 A/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A>Title: The complete genome sequence of *Escherichia coli* K-12.
 A/Reference number: A64720; MUID:97425617; PMID:9278503
 A/Accession: D64861
 A>Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-277 <BALT>
 A/Cross-references: GB:A5000215; GB:U00096; NID:91787405; PIDD:AACT4243.1; PID:91787406;
 A/Experimental source: strain K-12, substrain MG1655
 C/Genetics:
 A/Gene: mcrA; mcrJ; rglA
 A/Map position: 25 min
 C/Superfamily: Escherichia coli 5-methylcytosine-specific restriction enzyme A
 C/Keywords: endonuclease; hydrolase; restriction modification system

Query Match 66.0%; Score 31; DB 2; Length 277;
 Best Local Similarity 50.0%; Pred. No. 39;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 2 RXXHSMEXR 11
 Db 69 RRRMSLDER 78

RESULT 12
 552641
 heat shock transcription factor HSP1 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 A/Variety: columbica
 C/Date: 01-Aug-1995 #sequence_revision 24-Mar-1999 #text_change 16-Jun-2000
 C/Accession: F71447; S52641; S62227; S38873
 R/Beyran, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
 P.; Weller, H.; Wedler, E.; Wandt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
 arnagh, T.; Hempel, S.; Kotter, P.; Ertan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A/Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulidomenech
 erhof, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
 C.; Chaitzals, N.
 A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A/Reference number: A71400; MUID:96121113; PMID:9461215
 A/Accession: F71447
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-495 <BEV>
 A/Cross-references: GB:597344; NID:92245126; PIDD:CAI0555.1; PID:92245134
 R/Huebel, A.; Schaeffl, F.
 Plant Mol. Biol. 26, 353-362, 1994
 A>Title: Arabidopsis heat shock factor: isolation and characterization of the gene and t
 A/Reference number: 552641; MUID:95036006; PMID:7948861
 A/Accession: S52641
 A/Molecule type: DNA
 A/Residues: 1-215, 'M', 217-284, 'D', 286-335, 'T', 337-389, 'Y', 391-469, 474-495 <RUE>
 A/Cross-references: EMBL:X76167
 A/Accession: S62227
 A/Molecule type: mRNA
 A/Residues: 31-215, 'M', 217-284, 'D', 286-335, 'T', 337-389, 'Y', 391-469, 474-495 <HUF>
 R/Huebel, A.; Schaeffl, F.
 Submitted to the EMBL Data Library, November 1993
 A/Description: Arabidopsis heat shock factor: isolation and characterization of the gene
 A/Reference number: S38873
 A/Accession: S38873
 A/Molecule type: DNA

A/Residues: 1-215, 'M', 217-284, 'D', 286-335, 'T', 337-389, 'Y', 391-461, 474-495 <HUF>
 A/Cross-references: EMBL:X76167; NID:9429154; PIDD:CA55761.1; PID:9429155
 C/Genetics:
 A/Gene: HSP1
 A/Map position: 4COP9-4C3845
 A/Introns: 11/3
 A/Function: <YFC>
 A/Description: transcription factor that binds to heat shock promoter elements
 C/Superfamily: tomato heat shock transcription factor HSF8; HSF DNA-binding domain ho
 C/Keywords: DNA binding; heat shock; leucine zipper; nucleus; stress-induced protein;
 F/53-148/Domain: HSF DNA-binding domain homology <HSE>
 F/181-202/Region: leucine zipper

Query Match 66.0%; Score 31; DB 1; Length 495;
 Best Local Similarity 55.6%; Pred. No. 71;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 SRXHSME 9
 Db 277 SRRHSLE 285

RESULT 13
 H84075
 thioedoxin reductase BR3408 [Imported] - Bacillus halodurans (strain C-125)
 C/Species: Bacillus halodurans
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 03-Aug-2001
 C/Accession: H84075
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A/Reference number: A83650; MUID:20512582; PMID:11058132
 A/Accession: H84075
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-330 <STO>
 A/Cross-references: GB:AP001518; GB:BA000004; NID:910175792; PIDD:BAE07127.1; GSPDB:G
 A/Experimental source: strain C-125
 C/Genetics:
 A/Gene: BR3408
 C/Superfamily: thioedoxin reductase; thioedoxin reductase homology

Query Match 63.8%; Score 30; DB 2; Length 330;
 Best Local Similarity 62.5%; Pred. No. 75;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 RXXHSME 9
 Db 190 RAHESVE 197

RESULT 14
 B70015
 thioedoxin reductase homolog yunc - Bacillus subtilis
 C/Species: Bacillus subtilis
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C/Accession: B70015
 R/Kunzt, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emmeron, P.T.; Ertan, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henselt, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landino
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y., M.; Oawa, K.; Ogihara, A.; Oudega, B.; Park, S.B.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani
 A/Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowaka, A.; Se
 akuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, T.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A/Reference number: A69580; MUID:98040033; PMID:9384377

A:Accession: B70015

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-332 <KUN>

A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDD:CAD15201.1; PTD:g2635708

A:Experimental source: strain 168

C:Genetics:

A:Gene: yunc

C:Superfamily: thioresoxin reductase; thioresoxin reductase homology

Query Match

Best Local Similarity 63.8%; Score 30; DB 1; Length 332;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 RXHXSR 9

1 1 1 1 1

DB 190 RAHEHVE 197

RESULT 15

JC5166

ketol-acid reductoisomerase (EC 1.1.1.86) - Mycobacterium avium

N:Alternate names: acetoxyhydroxy acid isomerase

C:Species: Mycobacterium avium

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: JC5166

R:Gusberti, L.; Cantoni, R.; De Rossi, E.; Branzoni, M.; Riccardi, G.

Gene 177, 83-85, 1996

A>Title: Cloning and sequencing of the *llvnc* gene cluster from Mycobacterium avium.

A:Reference number: JC5164; MID:97080504; PMID:8921849

A:Accession: JC5166

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-332 <GUS>

A:Cross-references: GB:L49392; NID:g1196506; PIDD:AB38428.1; PTD:g1196509

C:Genetics:

A:Gene: llvnc

C:Superfamily: Methanococcus ketol-acid reductoisomerase; ketol-acid reductoisomerase hc;
C:Keywords: Intramolecular transferase; isoleucine-valine biosynthesis; isomerase; oxidid
F:18-200/Domain: ketol-acid reductoisomerase homology <KAR>

Query Match

Best Local Similarity 63.8%; Score 30; DB 1; Length 332;

Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXSR 11

1 1 1 1 1

DB 24 SGGHARSLR 34

Search completed: May 1, 2003, 14:31:42
Job time: 10.1875 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:33:06 ; Search time 6.875 Seconds

(without alignments)
138.059 Million cell updates/sec

Title: US-09-446-543a-74
Perfect score: 47
Sequence: 1 SRKHSHSMEXR 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PC1_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PC10_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	91.5	31	10	US-09-932-161-14
2	43	91.5	31	12	US-10-044-592-4
3	43	91.5	31	12	US-10-044-592-5
4	43	91.5	70	12	US-10-044-592-90
5	43	91.5	82	12	US-10-044-592-1
6	43	91.5	86	12	US-10-044-592-96
7	43	91.5	91	12	US-10-044-592-94
8	42	89.4	29	12	US-10-044-592-26
9	42	89.4	31	10	US-09-932-161-13
10	42	89.4	31	10	US-09-932-161-15
11	42	89.4	31	12	US-10-044-592-39
12	42	89.4	32	12	US-10-044-592-40
13	42	89.4	33	12	US-10-044-592-41
14	42	89.4	40	12	US-10-044-592-80
15	42	89.4	87	12	US-10-044-592-82
16	42	89.4	98	12	US-10-044-592-28
17	42	89.4	98	12	US-10-044-592-38
18	42	89.4	98	12	US-10-044-592-82
19	42	89.4	98	12	US-10-044-592-84

20	42	89.4	98	12	US-10-044-592-86	Sequence 86, App1
21	42	89.4	98	12	US-10-044-592-88	Sequence 88, App1
22	38	80.9	9	12	US-10-044-592-7	Sequence 7, App1
23	33	70.2	25	12	US-10-044-592-78	Sequence 78, App1
24	32	68.1	363	10	US-09-841-132-580	Sequence 580, App
25	31	66.0	103	10	US-09-764-877-1964	Sequence 1964, App
26	30	63.8	236	9	US-09-984-245-318	Sequence 318, App
27	30	63.8	236	9	US-09-966-262-318	Sequence 318, App
28	30	63.8	236	9	US-09-983-966-318	Sequence 318, App
29	30	63.8	236	9	US-10-143-090-318	Sequence 318, App
30	30	63.8	1224	10	US-09-801-368-222	Sequence 222, App
31	29	61.7	18	10	US-09-864-761-37659	Sequence 37659, A
32	29	61.7	525	10	US-09-804-5518-36	Sequence 36, App1
33	29	61.7	580	10	US-09-925-301-943	Sequence 943, App
34	29	61.7	774	9	US-09-974-298-122	Sequence 122, App
35	29	61.7	774	10	US-09-782-980-16	Sequence 16, App1
36	29	61.7	774	10	US-09-909-743-7	Sequence 7, App1
37	28	59.6	44	10	US-09-864-761-37970	Sequence 37970, A
38	28	59.6	113	10	US-09-916-790-27	Sequence 27, App1
39	28	59.6	338	9	US-09-738-626-6959	Sequence 6959, App
40	28	59.6	390	10	US-09-901-151-5	Sequence 5, App1
41	28	59.6	390	10	US-09-901-151-6	Sequence 6, App1
42	28	59.6	397	10	US-09-901-151-2	Sequence 2, App1
43	28	59.6	414	10	US-09-901-151-4	Sequence 4, App1
44	28	59.6	575	9	US-10-014-101-2	Sequence 2, App1
45	28	59.6	577	10	US-09-815-242-10193	Sequence 10193, A

ALIGNMENTS

```

RESULT 1
US-09-932-161-14
; Sequence 14, Application US/09932161
; Patent No. US20020037553A1
; GENERAL INFORMATION:
; APPLICANT: Civeilli, Olivier
; TITLE OF INVENTION: Screening and Therapeutic Methods For
; FILE REFERENCE: P-UC 4679
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US/09/932,161
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Rattus
US-09-932-161-14

Query Match      91.5%  Score 43;  DB 10;  Length 31;
Best Local Similarity 72.7%  Pred. No. 0.049;
Matches 8;  Conservative 0;  Mismatches 3;  Indels 0;  Gaps 0;

QY      1 SRKHSHSMEXR 11
DB      1 SRKHSHSMEXR 11
      |||||
      |||||

RESULT 2
US-10-044-592-4
; Sequence 4, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10

```

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO: 4
LENGTH: 31
TYPE: PRT
ORGANISM: Murine
US-10-044-592-4

Query Match 91.5%; Score 43; DB 12; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.049;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
||| |||||
Db 1 SRAHQSMETR 11

RESULT 3
US-10-044-592-5
Sequence 5, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO: 5
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(31)
OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 91.5%; Score 43; DB 12; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.049;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
||| |||||
Db 1 SRAHQSMETR 11

RESULT 4
US-10-044-592-90
Sequence 90, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639

PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO: 90
LENGTH: 70
TYPE: PRT
ORGANISM: Rat
US-10-044-592-90

Query Match 91.5%; Score 43; DB 12; Length 70;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 14
||| |||||
Db 22 SRAHQSMETR 32

RESULT 5
US-10-044-592-1
Sequence 1, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO: 1
LENGTH: 82
TYPE: PRT
ORGANISM: Murine
US-10-044-592-1

Query Match 91.5%; Score 43; DB 12; Length 82;
Best Local Similarity 72.7%; Pred. No. 0.13;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
||| |||||
Db 21 SRAHQSMETR 31

RESULT 6
US-10-044-592-96
Sequence 96, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28

NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 96
LENGTH: 86
TYPE: PRT
ORGANISM: mammalian
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1224)..(1243)
OTHER INFORMATION: Bracket region depicted in FIG 39.
US-10-044-592-96

Query Match 91.5%; Score 43; DB 12; Length 86;
Best Local Similarity 72.7%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 22 SRAHSHMETR 32

RESULT 7
US-10-044-592-94
Sequence 94, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 94
LENGTH: 91
TYPE: PRT
ORGANISM: Mammalian
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(31)
OTHER INFORMATION: primer
NAME/KEY: misc.feature
LOCATION: (925)..(955)
OTHER INFORMATION: primer
NAME/KEY: misc.feature
LOCATION: (1)..(955)
OTHER INFORMATION: Insert fragment of pmgB3 as depicted in FIG 36 and 37.
US-10-044-592-94

Query Match 91.5%; Score 43; DB 12; Length 91;
Best Local Similarity 72.7%; Pred. No. 0.15;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 22 SRAHSHMETR 32

RESULT 8
US-10-044-592-26
Sequence 26, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use

FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 26
LENGTH: 29
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-26

Query Match 89.4%; Score 42; DB 12; Length 29;
Best Local Similarity 72.7%; Pred. No. 0.071;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 1 SRAHSHMETR 11

RESULT 9
US-09-932-161-13
Sequence 13, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Civealli, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods for
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 31
TYPE: PRT
ORGANISM: Bos taurus
US-09-932-161-13

Query Match 89.4%; Score 42; DB 10; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 1 SRAHSHMETR 11

RESULT 10
US-09-932-161-15
Sequence 15, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Civealli, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods for
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 15
LENGTH: 31
TYPE: PRT
ORGANISM: Homo Sapien
US-09-932-161-15

Query Match 89.4%; Score 42; DB 10; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 STHRHSMETR 11

RESULT 11
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 89.4%; Score 42; DB 12; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQSMETR 11

RESULT 12
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 40
LENGTH: 32
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-40

Query Match 89.4%; Score 42; DB 12; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.079;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQSMETR 11

RESULT 13
US-10-044-592-41
Sequence 41, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 41
LENGTH: 33
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-41

Query Match 89.4%; Score 42; DB 12; Length 33;
Best Local Similarity 72.7%; Pred. No. 0.082;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQSMETR 11

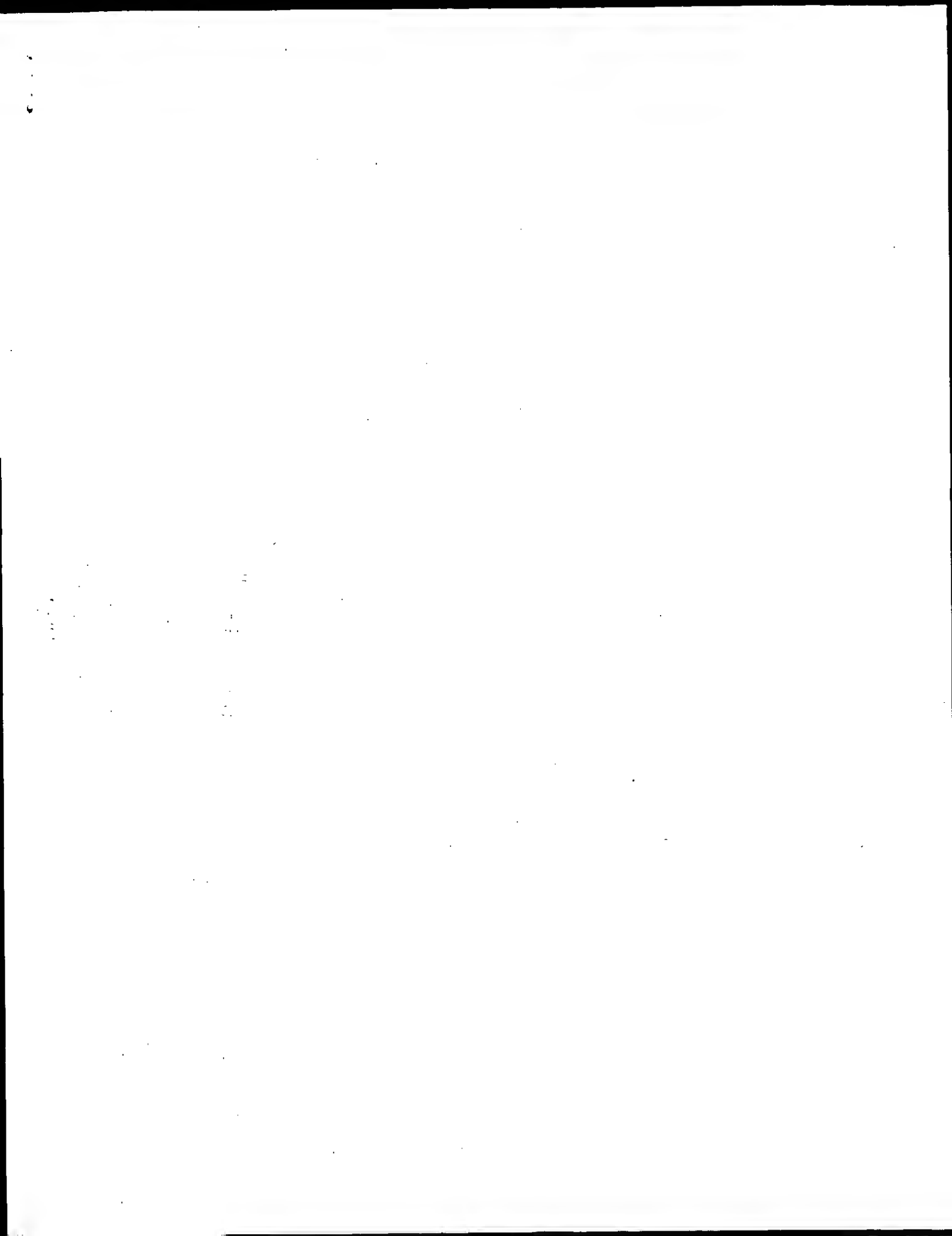
RESULT 14
US-10-044-592-80
Sequence 80, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 80
LENGTH: 40
TYPE: PRT
ORGANISM: Bovine
FEATURE:
NAME/KEY: misc-feature
LOCATION: (99)-(126)
OTHER INFORMATION: primer
US-10-044-592-80

Query Match 89.4%; Score 42; DB 12; Length 40;
 Best Local Similarity 72.7%; Pred. No. 0.099;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 ||| ||| |
 Db 23 SRXHXSMEXR 33

RESULT 15
 US-10-044-592-92
 ; Sequence 92, Application US/10044592
 ; Patent No. US20020143152A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinuma, Shuji
 ; APPLICANT: Fukuoka, Shoji
 ; TITLE OF INVENTION: Polypeptides, their production and use
 ; FILE REFERENCE: 2463052P
 ; CURRENT APPLICATION NUMBER: US/10/044,592
 ; CURRENT FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 09/403639
 ; PRIOR FILING DATE: 1999-25-10
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01923
 ; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-109974
 ; PRIOR FILING DATE: 1997-04-28
 ; NUMBER OF SEQ ID NOS: 96
 ; SOFTWARE:
 ; SEQ ID NO 92
 ; LENGTH: 87
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-044-592-92

Query Match 89.4%; Score 42; DB 12; Length 87;
 Best Local Similarity 72.7%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 ||| ||| |
 Db 23 SRXHXSMEXR 33

Search Completed: May 1, 2003, 14:46:07
 Job time : 7.875 secs



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:13 ; Search time 5.32812 Seconds

(without alignments)
60.744 Million cell updates/sec

Title: US-09-446-543a-74
Sequence: 1 SRXHXSMEXR 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/1aa/5a_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5b_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5a_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5b_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCUTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/ackfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	91.5	31	3	US-09-105-678a-8
2	43	91.5	31	3	US-09-105-678a-37
3	43	91.5	31	4	US-09-172-353-4
4	43	91.5	31	4	US-08-776-971-47
5	43	91.5	31	4	US-09-421-208-8
6	43	91.5	31	4	US-09-421-208-37
7	43	91.5	31	4	US-09-560-915-14
8	43	91.5	32	3	US-09-105-678a-38
9	43	91.5	32	3	US-08-776-971-48
10	43	91.5	32	3	US-09-421-208-38
11	43	91.5	33	3	US-09-105-678a-39
12	43	91.5	33	4	US-08-776-971-49
13	43	91.5	33	4	US-08-776-971-45
14	43	91.5	83	4	US-08-776-971-124
15	43	91.5	83	4	US-08-776-971-137
16	43	91.5	15	4	US-08-776-971-92
17	42	89.4	29	4	US-09-105-678a-29
18	42	89.4	29	4	US-08-776-971-3
19	42	89.4	29	4	US-09-421-208-29
20	42	89.4	31	3	US-09-105-678a-7
21	42	89.4	31	3	US-09-105-678a-9
22	42	89.4	31	3	US-09-105-678a-31
23	42	89.4	31	3	US-09-105-678a-43
24	42	89.4	31	4	US-08-776-971-61
25	42	89.4	31	4	US-08-776-971-97
26	42	89.4	31	4	US-08-776-971-97
27	42	89.4	31	4	US-08-776-971-97

28	42	89.4	31	4	US-09-421-208-7	Sequence 7, Appl
29	42	89.4	31	4	US-09-421-208-9	Sequence 9, Appl
30	42	89.4	31	4	US-09-421-208-31	Sequence 31, Appl
31	42	89.4	31	4	US-09-421-208-43	Sequence 43, Appl
32	42	89.4	31	4	US-09-560-915-15	Sequence 15, Appl
33	42	89.4	31	4	US-09-560-915-13	Sequence 13, Appl
34	42	89.4	32	3	US-09-105-678a-32	Sequence 32, Appl
35	42	89.4	32	3	US-09-105-678a-44	Sequence 44, Appl
36	42	89.4	32	4	US-08-776-971-62	Sequence 62, Appl
37	42	89.4	32	4	US-09-421-208-32	Sequence 42, Appl
38	42	89.4	32	4	US-09-421-208-44	Sequence 44, Appl
39	42	89.4	33	3	US-09-105-678a-33	Sequence 33, Appl
40	42	89.4	33	3	US-09-105-678a-45	Sequence 45, Appl
41	42	89.4	33	4	US-08-776-971-7	Sequence 7, Appl
42	42	89.4	33	4	US-08-776-971-63	Sequence 63, Appl
43	42	89.4	33	4	US-09-421-208-33	Sequence 33, Appl
44	42	89.4	33	4	US-09-421-208-45	Sequence 45, Appl
45	42	89.4	33	4	US-09-421-208-45	Sequence 45, Appl

ALIGNMENTS

Result 1
US-09-105-678a-8
Sequence 8, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678a-8
Query Match 91.5%, Score 43, DB 3, Length 31,
Best Local Similarity 72.7%, Pred. No. 0.013,
Matches 8, Conservative 0, Mismatches 3, Indels 0,
Gaps 0,
QY 1 SRXHXSMEXR 11

Db 1 SRHXHSMETR 11

RESULT 2
US-09-105-678A-37

Sequence 37, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Sueaaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-37

Query Match

Best Local Similarity 91.5%; Score 43; DB 3; Length 31;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMETR 11

Db 1 SRHXHSMETR 11

RESULT 3

US-09-172-353-4

Sequence 4, Application US/09172353

Patent No. 6197530

GENERAL INFORMATION:

APPLICANT: Stricker-Kongra, Alain

TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS

FILE REFERENCE: 07334/102001

CURRENT APPLICATION NUMBER: US/09/172,353

CURRENT FILING DATE: 1998-10-14

NUMBER OF SEQ ID NOS: 7

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 31

TYPE: PRT
ORGANISM: Mus musculus
US-09-172-353-4

Query Match

Best Local Similarity 91.5%; Score 43; DB 4; Length 31;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMETR 11

Db 1 SRHXHSMETR 11

RESULT 4

US-08-776-971-47

Sequence 47, Application US/08776971B

Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hiduma, Shuji

APPLICANT: Habata, Yugo

APPLICANT: Kawamata, Yuji

APPLICANT: Hosoya, Masaki

APPLICANT: Fujii, Ryo

APPLICANT: Fukusumi, Shoji

APPLICANT: Kitada, Chieko

TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-08-776-971-47

Query Match

91.5%; Score 43; DB 4; Length 31;

Best Local Similarity 72.7%; Pred. No. 0.013; 3; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11
DB 1 SRAHSHMETR 11

RESULT 5

US-09-421-208-8
Sequence 8, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-8

Query Match 91.5%; Score 43; DB 4; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11
DB 1 SRAHSHMETR 11

RESULT 6

US-09-421-208-37
Sequence 37, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-37

Query Match 91.5%; Score 43; DB 4; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11
DB 1 SRAHSHMETR 11

RESULT 7

US-09-560-915-14
Sequence 14, Application US/09560915
Patent No. 6383764
GENERAL INFORMATION:
APPLICANT: Clivell, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Therapeutic Compositions and Methods
TITLE OF INVENTION: Relating to Prolactin Releasing Peptide (Prnp)
FILE REFERENCE: P-DC 3534
CURRENT APPLICATION NUMBER: US/09/560,915
CURRENT FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 31
TYPE: PRT
ORGANISM: Rattus
US-09-560-915-14

Query Match 91.5%; Score 43; DB 4; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.013; 3; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11
DB 1 SRAHSHMETR 11

DB 1 SRAHQSMETR 11

RESULT 8

US-09-105-678A-38

Sequence 38, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conltd, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-38

QY 1 SRXHXSMETR 11
DB 1 SRAHQSMETR 11Query Match 91.5%; Score 43; DB 3; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.014;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 9

US-08-776-971-48

Sequence 48, Application US/08776971B
Patent No. 6226984

GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji

APPLICANT: Hadata, Yugo

APPLICANT: Kawamata, Yui

APPLICANT: Hosoya, Masaki

APPLICANT: Fujii, Ryo

APPLICANT: Fukusumi, Shoji

APPLICANT: Kitade, Chieko

TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESS:

ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSD for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03621

FILING DATE: 28-Dec-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-Dec-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conltd, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 48:

QY 1 SRXHXSMETR 11
DB 1 SRAHQSMETR 11Query Match 91.5%; Score 43; DB 4; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.014;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 10

US-09-421-208-38

Sequence 38, Application US/09421208
Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-38

Query Match 91.5%; Score 43; DB 4; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.014;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11
II IIII I
DB I SRAHQSMETR 11

RESULT 11
US-09-105-678a-39
Sequence 39, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-6440
TELEFAX: 617-523-3400
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:

LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678a-39

Query Match 91.5%; Score 43; DB 3; Length 33;
Best Local Similarity 72.7%; Pred. No. 0.014;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11
II IIII I
DB 1 SRAHQSMETR 11

RESULT 12
US-08-776-971-49
Sequence 49, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hanuma, Shuji
APPLICANT: Habata, Yugo
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
SEQUENCE DESCRIPTION: SEQ ID NO: 49:

US-08-776-971-49

Query Match	91.58;	Score 43;	DB 4;	Length 33;
Best Local Similarity	72.76;	Pred. No. 0.014;		
Matches	8; Conservative	0; Mismatches	3; Indels	

```

OY      1 SRHXHSMEXR 11
        11 1 1111 1
Db      1 SRAHOSMETR 11

```

RESULT 13
US-09-421-208-39

Sequence 39, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston

Query Match	91.5%	Score 43;	DB 4;	Length 33;
Best Local Similarity	72.7%	Pred. No. 0.014;		
Matches	8;	Conservative	0;	Mismatches 3; Indels 0;

```

QY      1 SRXHXHSMEXR 11
        11 111111
Db      1 SRAHQSMETR 11

```

RESULT 14
US-08-776-971-45
: Sequence 45, Application US/08776971B
: Patent No. 6228984
: GENERAL INFORMATION:
: APPLICANT: Rituuma, Shuji

Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Sho-ji
Kitada, Chieko

TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS: 140

Query Match	91.5%	Score 43;	DB 4;	Length 83;
Best Local Similarity	72.7%	Pred. No. 0.036;		
Matches	8;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;

```

QY      1 SRXHXHSMEXR 11
          ||| ||| |
Db      22 SRAHQHSMETR 32

```

RESUL 15
US-08-776-971-124
Sequence 124, Application US/0876971A
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hiruma, Shuji
Hadamata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Fukushima, Shoji

Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776, 971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: COLLIN, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 124:
SEQUENCE CHARACTERISTICS:
LENGTH: 83 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 124:
US-08-776-971-124
Query Match 91.5%; Score 43; DB 4; Length 83;
Best Local Similarity 72.7%; Pred. No. 0.036;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRKHSHSMEXR 11
||| |||||
Db 22 SRAHQSHSMETR 32

Search completed: May 1, 2003, 14:33:54
Job time: 5.32812 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 24.0625 Seconds
(without alignments)
188.386 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDINPAMYXKRGIRPVGRFX 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VIRTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	95.4	98	6 Q8WN12	Q8WN12 oviss aries
2	82	75.2	117	13 Q9W624	Q9W624 carassius a
3	49.5	45.4	664	16 Q911W4	Q911W4 pseudomonas
4	49	45.0	790	10 Q9M371	Q9M371 arabisdopsis
5	46.5	42.7	420	2 Q931Z7	Q931Z7 streptomyces
6	46	42.2	54	4 Q90UF9	Q90UF9 homo sapien
7	46	42.2	333	16 Q9PH76	Q9PH76 xyloella fas
8	46	42.2	465	4 Q8W885	Q8W885 homo sapien
9	46	42.2	540	10 Q9L6Z0	Q9L6Z0 oryza sativ
10	46	42.2	540	10 Q9L6Z0	Q9L6Z0 caenorhabdi
11	45	41.3	767	5 Q952B5	Q952B5 leishmania
12	44	40.4	145	5 Q952B5	Q952B5 pseudomonas
13	44	40.4	250	16 Q912F2	Q912F2 pyrococcus
14	44	40.4	284	17 Q8U112	Q8U112 pyrococcus
15	44	40.4	284	17 Q8U112	Q8U112 pyrococcus
16	44	40.4	419	4 Q9Y276	Q9Y276 homo sapien

17	44	40.4	884	15 Q8Q718	Q8Q718 human immun
18	43.5	39.9	1501	10 Q9SD86	Q9SD86 arabisdopsis
19	43	39.4	220	16 Q91022	Q91022 pseudomonas
20	43	39.4	226	16 Q8RF10	Q8RF10 fusobacteri
21	43	39.4	232	16 Q9R221	Q9R221 deinococcus
22	43	39.4	309	5 Q17234	Q17234 caenorhabdi
23	43	39.4	333	4 Q96SD4	Q96SD4 homo sapien
24	43	39.4	370	11 Q9D3V7	Q9D3V7 mus musculu
25	43	39.4	418	11 Q9CAZ5	Q9CAZ5 mus musculu
26	43	39.4	476	16 Q8XA75	Q8XA75 escherichia
27	43	39.4	813	17 Q8ZXH2	Q8ZXH2 pyrobaculum
28	43	39.4	1488	5 Q20294	Q20294 caenorhabdi
29	42.5	39.0	333	16 Q9KJ10	Q9KJ10 streptomyces
30	42.5	39.0	360	5 Q19879	Q19879 caenorhabdi
31	42	38.5	293	2 Q53855	Q53855 spiroplasma
32	42	38.5	416	16 Q99Z7A	Q99Z7A streptococ
33	42	38.5	428	16 Q9PKF7	Q9PKF7 chlamydia m
34	42	38.5	433	16 Q9A382	Q9A382 caulobacter
35	42	38.5	467	11 Q8R054	Q8R054 mus musculu
36	42	38.5	549	2 Q9EW96	Q9EW96 streptomyces
37	41.5	38.1	151	12 Q91R01	Q91R01 dashen mos
38	41	37.6	97	2 Q33440	Q33440 pseudomonas
39	41	37.6	128	16 P74747	P74747 syncytocyst
40	41	37.6	226	16 Q9A359	Q9A359 caulobacter
41	41	37.6	263	16 Q8XRE9	Q8XRE9 ralsstonia s
42	41	37.6	273	17 Q8ZTH7	Q8ZTH7 pyrobaculum
43	41	37.6	280	4 Q969Z5	Q969Z5 homo sapien
44	41	37.6	284	17 Q9V003	Q9V003 pyrococcus
45	41	37.6	328	4 Q9NVR5	Q9NVR5 homo sapien

ALIGNMENTS

RESULT 1

ID Q8WN12 PRELIMINARY; PRT; 98 AA.
AC Q8WN12
DI 01-MAR-2002 (TREMBLrel. 20, Created)
DI 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DI 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Preprolactin-releasing peptide.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Curlewis J.D., Kusters D.H.L., Barclay J.L., Anderson S.T.;
RT "Prolactin-releasing peptide (PRP) in the ewe: cDNA cloning, mRNA
RT distribution and effects on prolactin secretion in vitro and in
RT vivo".
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF450453; AAL47178.1; ..2A53331EP62CAAB5 CRC64;
SQ SEQUENCE 98 AA; 10513 MW; 2A53331EP62CAAB5 CRC64;

Query Match 95.4%; Score 104; DB 6; Length 98;
Best Local Similarity 90.0%; Pred. No. 1.2e-05;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TPDINPAMYXKRGIRPVGRF 20
DB 34 TPDINPAMYXKRGIRPVGRF 53
RESULT 2
Q9W624 PRELIMINARY; PRT; 117 AA.
AC Q9W624
DI 01-NOV-1999 (TREMBLrel. 12, Created)
DI 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DI 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

GN D049J7.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lawlor S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035608; CAB55682.1;
 FT NON_TER 54 54
 SO SEQUENCE 54 AA; 6110 MW; E2F3C39F7B961A9F CRC64;

Query Match 42.2%; Score 46; DB 4; Length 54;
 Best Local Similarity 43.8%; Pred. No. 2.4;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPDINPANYXXRGIRP 16
 18 TPVPTWYAGSGYYP 33

RESULT 7
 O9PH76 PRELIMINARY; PRT; 333 AA.
 AC O9PH76;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hydroxybenzoate octaprenyltransferase.
 GN XF0068
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Lemos M.A., Madalena A.M.B.N., Madalena H.M.F., Martino C.L.,
 RA Marques M.V., Martins A.A.B., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montelero-Vitorillo C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.R.S.,
 RA Nhami A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zait M., Zeldanis J., Zetahal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL; AE003860; AAF2881.1;
 DR InterPro; IPR000537; UBLA.
 DR Pfam; PF01040; UBLA.1.
 DR PROSITE; PS00943; UBLA; UNKNOWN.1.
 KW Complete proteome.
 SO SEQUENCE 333 AA; 37931 MW; ECF3F4716C962B95 CRC64;

Query Match 42.2%; Score 46; DB 16; Length 333;
 Best Local Similarity 53.3%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 4 INPANYXXRGIRPVG 18
 54 IDPWTYKARGDRPVG 68

RESULT 8
 O60687 PRELIMINARY; PRT; 465 AA.
 AC O60687;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sush1-repeat protein.
 GN SRPUL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kurosawa H., Inukai T., Inaba T., Gol K., Chang K.-S., Sinjo T.,
 RA Rakesstraw K.M., Naeve C.W., Look T.A.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF060567; AAC15765.1;
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR003410; HyalIn.
 DR InterPro; IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF02494; HVR.1.
 DR Pfam; PF00084; sush1.3.
 DR SMART; SM00032; CCP.3.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN.1.
 SO SEQUENCE 465 AA; 52971 MW; 4D752B187FE3EEB8 CRC64;

Query Match 42.2%; Score 46; DB 4; Length 465;
 Best Local Similarity 43.8%; Pred. No. 23;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPDINPANYXXRGIRP 16
 18 TPVPTWYAGSGYYP 33

RESULT 9
 O8W85 PRELIMINARY; PRT; 465 AA.
 AC O8W85;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sush1-repeat protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Strauberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC020733; AAR20733.1;
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR003410; HyalIn.
 DR InterPro; IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF02494; HVR.1.
 DR Pfam; PF00084; sush1.3.
 DR SMART; SM00032; CCP.3.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN.1.
 SO SEQUENCE 465 AA; 52937 MW; 3D7229487DA1B8BD CRC64;

Query Match 42.2%; Score 46; DB 4; Length 465;
 Best Local Similarity 43.8%; Pred. No. 23;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRP 16
 ||:|||||
 DB 18 TPATPTWYAGSGYRP 33

RESULT 10

OY 091GZ0 PRELIMINARY; PRT; 540 AA.
 AC 091GZ0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE EST AU070209(R3722) corresponds to a region of the predicted
 DE gene.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV, NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:PO702F03.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC PROTEINS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +
 CC DIPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.
 CC -1- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
 CC -1- METABOLISM: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
 CC THIOLESTER FORMATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
 DR EMBL: AP002481; BAA96583.1; -;
 DR HSSP: P06104; IAYZ.
 DR InterPro: IPR000608; UBQ_conjugat.
 DR Pfam: PF00179; UBQ_con; 1.
 DR ProDom: PD000461; UBQ_conjugat; 1.
 DR SMART: SM00212; UBQC; 1.
 DR PROSITE: PS50127; UBIQUITIN_CONJUGAT_2; 1.
 KM Ligase; ubiquitin conjugation.
 SQ SEQUENCE 540 AA; 60487 MW; 5DE1FF4EB75A86E CRC64;

Query Match 42.2%; Score 46; DB 10; Length 540;
 Best Local Similarity 50.0%; Pred. No. 27;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 DINPAMYXXRGIRP 16
 ||:|||||
 DB 69 DLGYAMWVRKGLRP 82

RESULT 11

OY 020170 PRELIMINARY; PRT; 767 AA.
 AC 020170;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE F38E11.7 protein.
 GN F38E11.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RX [1]
 RN SEQUENCE FROM N.A.

RA Matthews P.;
 RN Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;

RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z68342; GAA92775.1; -;
 DR InterPro: IPR000595; cNMP_binding.
 DR InterPro: IPR001622; K-channel_pore.
 DR InterPro: IPR000636; K-channel_nlg.
 DR Pfam: PF00027; cNMP_binding; 1.
 DR Pfam: PF00520; Ion_trans; 1.
 DR SMART: SM00100; cNMP; 1.
 DR PROSITE: PS00888; cNMP_BINDING_1; UNKNOWN_1.
 DR PROSITE: PS50042; cNMP_BINDING_3; 1.
 SQ SEQUENCE 767 AA; 89988 MW; F7ECF69DBEACF3 CRC64;

Query Match 41.3%; Score 45; DB 5; Length 767;
 Best Local Similarity 75.0%; Pred. No. 57;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 PDINPAMY 9
 ||:|||||
 DB 747 PDYKPAWY 754

RESULT 12

OY 095ZB5 PRELIMINARY; PRT; 145 AA.
 AC 095ZB5;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 15.8 kDa protein.
 GN L1994.09.
 OS Leishmania major.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=5664;
 RX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=FRIDLIN;
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Rieger M.,
 RA Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIDLIN;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome."
 RL Genome Res. 8:135-145(1998).
 DR EMBL: AL596272; CAC44727.1; -;
 KM Hypothetical protein.
 SQ SEQUENCE 145 AA; 15806 MW; F30454BC29984F83 CRC64;

Query Match 40.4%; Score 44; DB 5; Length 145;
 Best Local Similarity 40.0%; Pred. No. 15;
 Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRP 20
 ||:|||||
 DB 6 TPANPALALRVHPARM 25

RESULT 13

OY 0912F2 PRELIMINARY; PRT; 250 AA.
 AC 0912F2;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
 DE Hypothetical protein PA1952.
 GN PA1952.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas;
 NCBI_TaxID=287;
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Huinagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardy K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen."
 RL Nature 406:959-964(2000).
 DR EMBL; AF004622; AAG05340.1;
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 250 AA; 25619 MW; B997F6B28D792C2 CRC64;

Query Match 40.4%; Score 44; DB 16; Length 250;
 Best local similarity 47.1%; Pred. No. 26;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRPVGR 18
 DB 233 PALNCAFEQIALRPSG 249

RESULT 14
 ID 050128 PRELIMINARY; PRT; 284 AA.
 AC 050128;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
 DE Hypothetical protein PH1420.
 GN PH1420.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawarayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
 Yamamoto B., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 Sekai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 Aoki K.-I., Yoshitawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
 RL DNA Res. 5:55-76(1998).
 DR EMBL; AF000006; BAA30526.1;
 DR InterPro; IPR001330; Prenyltrans.
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00432; Prenyltrans; 2.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 284 AA; 32319 MW; 8E0E7BC3711D3815 CRC64;

Query Match 40.4%; Score 44; DB 17; Length 284;
 Best local similarity 44.4%; Pred. No. 30;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRPVGR 19

DB 217 PYTEPFYALRGLELLGR 234

RESULT 15
 ID 080112 PRELIMINARY; PRT; 284 AA.
 AC 080112;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Hypothetical protein PF1418.
 GN PF1418.
 OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome."
 RL Submitted (FEF-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF010244; AAL81542.1;
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 284 AA; 32302 MW; ID4C5746A1831390 CRC64;

Query Match 40.4%; Score 44; DB 17; Length 284;
 Best local similarity 44.4%; Pred. No. 30;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRPVGR 19
 DB 217 PYTEPFYALRGLELLGR 234

Search completed: May 1, 2003, 14:32:57
 Job time: 28.0625 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12, Search time 12.375 Seconds
(without alignments)
170.906 Million cell updates/sec

Title: US-09-446-543a-73
Perfect score: 109
Sequence: 1 TPDINPAWYXXRGIRPVGRFX 22

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR:***
2: PIR:***
3: PIR:***
4: PIR:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	95.4	83	JC7607	prolactin-releasing
2	49.5	45.4	664	F83376	conserved hypothet
3	49	45.0	790	T47959	hypothetical prote
4	46	42.2	333	H82852	hydroxybenzoate oc
5	45	41.3	767	T21969	hypothetical prote
6	44	40.4	250	G83400	hypothetical prote
7	44	40.4	284	F71015	hypothetical prote
8	43.5	39.9	1501	T43623	hypothetical prote
9	43	39.4	220	C83392	probable glutathio
10	43	39.4	232	G75608	hypothetical prote
11	43	39.4	309	T32376	hypothetical prote
12	43	39.4	376	A48197	opsin, ocellar
13	43	39.4	376	B48197	opsin, lateral eye
14	43	39.4	476	G64720	probable amino aci
15	43	39.4	476	G90629	probable inner mem
16	43	39.4	476	G85480	inner membrane tra
17	43	39.4	719	S61046	ARPI protein - yea
18	43	39.4	1296	T16859	hypothetical prote
19	42.5	39.0	443	T21499	hypothetical prote
20	42	38.5	428	F81694	hypothetical prote
21	42	38.5	433	H87660	private dehydroge
22	41.5	38.1	345	D84012	peptidoglycan-bind
23	41	37.6	128	S76955	hypothetical prote
24	41	37.6	226	A87664	hypothetical prote
25	41	37.6	284	A75117	hypothetical prote
26	41	37.6	338	T20100	hypothetical prote
27	41	37.6	342	B64395	malic acid transpo
28	41	37.6	343	T46534	probable FMN-depen
29	41	37.6	347	H64371	malic acid transpo

30	41	37.6	419	2	AH3166
31	41	37.6	476	2	AG0502
32	41	37.6	545	2	A87448
33	41	37.6	986	1	OTURGA
34	41	37.6	2155	2	T30197
35	40.5	37.2	779	2	T49717
36	40.5	37.2	1540	2	T45619
37	40	36.7	184	2	T35841
38	40	36.7	324	2	T35901
39	40	36.7	329	2	H70744
40	40	36.7	341	2	T35426
41	40	36.7	359	2	T40084
42	40	36.7	390	2	G82844
43	40	36.7	430	1	B69009
44	40	36.7	462	2	T00708
45	40	36.7	468	2	C83160

ALIGNMENTS

RESULT 1

JC7607 prolactin-releasing peptide - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001

C/Accession: JC7607

Riyamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, Blochm. Biophys. Res. Commun. 281, 53-56, 2001

A>Title: Isolation and characterization of the rat prolactin-releasing peptide gene:

A/Reference number: JC7607; MUID:21092785; PMID:11178959

A/Contents: Spleen

A/Accession: JC7607

A/Molecule type: DNA

A/Residues: 1-83 <YAM>

C/Cross-references: DDBJ:AB040612; DDBJ:AB040613

C/Comment: This peptide induces arachidonic acid metabolite release from rat anterior release, and stimulation of ACTH secretion from the pituitary.

C/Genetics:

A/Genes: PRRP

A/Introns: 33/1

Query Match

Best Local Similarity 95.4%; Score 104; DB 2; Length 83;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAWYXXRGIRPVGRF 20

Db 33 TPDINPAWYXXRGIRPVGRF 52

RESULT 2

F83376 conserved hypothetical protein PA2151 [Imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C/Accession: F83376

RiStover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A/Reference number: AB2950; MUID:20437337; PMID:10984043

A/Accession: F83376

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-664 <STO>

A/Cross-references: GB:AE004642; GB:AE004091; MUID:g9948163; PIDD:AGC05539.1; GSPDB:GN

A/Experimental source: strain PA01

A/Genes: PA2151

Query Match 45.4%; Score 49.5; DB 2; Length 664;

hypothetical prote
probable amino-act
conserved hypothet
spect receptor p
alpha tectorin - m
related to PCS1 pr
hypothetical prote
probable membrane
probable arac fami
hypothetical prote
probable oxidoredu
PWWP domain protei
cysteine synthase
conserved hypothet
violaxanthin de-ep
nitrile extrusion

Best Local Similarity 62.5%; Pred. No. 3.8;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 TPDINPAMYXXRGIRP 16
Db 478 TPDINP-WFLQRCRGP 492

RESULT 3

T47959

hypothetical protein F15G16.60 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47959

R:De Ban, M.; Maare, A.C.; Griwall, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet submitted to the Protein Sequence Database, January 2000

A:Reference number: Z24480

A:Accession: T47959

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-790 <DEH>

A:Cross-references: EMBL:AL132959

A:Experimental source: cultivar Columbia; BAC clone F15G16

C:Genetics:

A:Map position: 3

A:Introns: 39/1; 678/2; 698/3; 773/2

A:Note: F15G16.60

Query Match 45.0%; Score 49; DB 2; Length 790;
Best Local Similarity 47.4%; Pred. No. 5.5;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 PDINPAMYXXRGIRP 20
Db 366 PPHNPRRTGSGRLQPRGM 384

RESULT 4

H82852

hydroxybenzoate octaprenyltransferase XF0068 [Imported] - Xylella fastidiosa (strain 9a5

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000

C:Accession: H82852

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:2035717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82852

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-333 <STN>

A:Cross-references: GB:AE003860; GB:AE003849; NID:g9104830; PIDN:AAF82881.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Stimpson, A.J.G.; Reinsch, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; From

J.D.; Jungueira, M.L.; Kemper, E.L.; Kitejima, J.P.; Krieger, J.E.; Kurama, E.R.; Laig

chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaiah

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0068

C:Superfamily: 4-hydroxybenzoate octaprenyltransferase

Query Match

42.2%; Score 46; DB 2; Length 333;

Best Local Similarity 53.3%; Pred. No. 7.3;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 INPAMYXXRGIRP 18
Db 54 IDPWKRLAKGRDPRG 68

RESULT 5

T21969

hypothetical protein F38E11.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21969

R:Matthews, P. submitted to the EMBL Data Library, January 1996

A:Reference number: Z19495

A:Accession: T21969

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-767 <MTL>

A:Cross-references: EMBL:Z68342; PIDN:CAA92775.1; GSPDB:GN000022; CESP:F38E11.7

A:Experimental source: clone F38E11

C:Genetics:

A:Map position: 4

A:Introns: 50/2; 118/1; 139/2; 189/3; 226/1; 248/1; 287/2; 375/2; 432/3; 465/3; 548/1

Query Match 41.3%; Score 45; DB 2; Length 767;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PDINPAMY 9
Db 747 PDYKPRMY 754

RESULT 6

G83400

hypothetical protein PA1952 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83400

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

Nature 406, 959-964, 2000

A:Reference number: A82950; PMID:20437337; PMID:10984043

A:Accession: G83400

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-250 <STO>

A:Cross-references: GB:AE004622; GB:AE004091; NID:g9947948; PIDN:AG05340.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1952

Query Match 40.4%; Score 44; DB 2; Length 250;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 PDINPAMYXXRGIRP 18
Db 233 PALNCAMWGLRALRPSG 249

RESULT 7

F71015

hypothetical protein PH1420 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: F71015

R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Seki, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, DNA Res. 5, 53-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
 A:Reference number: A71000; MUID:98344137; PMID:9679194
 A:Accession: F71015
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-284 <RAW>
 A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30526.1; PID:g3257843
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH420
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH420

Query Match 40.4%; Score 44; DB 2; Length 284;
 Best Local Similarity 44.4%; Pred. No. 13;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 Oy 2 PDINPAMYXXRGIRPVR 19
 Db 217 PYRPFYALRGLELGR 234

RESULT 8

hypothetical protein F13G24.180 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T45623
 R:Bevan, M.; Van der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Be submitted to the Protein Sequence Database, December 1999
 A:Reference number: Z23009
 A:Accession: T45623
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1501 <BEV>
 A:Cross-references: EMBL:AL133421
 A:Experimental source: cultivar Columbia; BAC clone F13G24
 C:Genetics:
 A:Map position: 5
 A:Insertions: 64/1; 739/3; 785/2; 1302/2; 1318/3; 1399/2; 1434/2
 A:Note: F13G24.180

Query Match 39.9%; Score 43.5; DB 2; Length 1501;
 Best Local Similarity 34.5%; Pred. No. 84;
 Matches 10; Conservative 5; Mismatches 5; Indels 9; Gaps 2;
 Oy 1 TPDINPAMYXXRG-----IRPV---GRF 20
 Db 1184 SPQMAPSWISQYCTFKNGLVQPVNDGRF 1212

RESULT 9

C83292

Probable glutathione B-transferase PA2821 [Imported] - Pseudomonas aeruginosa (strain P
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Mar-2001
 C:Accession: C83292
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lim,
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: C83292
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-220 <STO>
 A:Cross-references: GB:AE004709; GB:AE004091; NID:g9948904; PIDN:AAG06209.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:

A:Gene: PA2821
 C:Superfamily: plase glutathione transferase

Query Match 39.4%; Score 43; DB 2; Length 220;
 Best Local Similarity 64.3%; Pred. No. 15;
 Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;
 Oy 6 PAMYXXRGIRPVR 19
 Db 39 PAMY--REISPLGR 50

RESULT 10

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: G75608
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Bickey, E.K.; Peterson, J.D.; Dodson, R.J
 S.; Smith, R.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radiotolerant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: G75608
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-232 <WHI>
 A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAE12317.1; PID:g646
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DRA0132
 A:Map position: 2

Query Match 39.4%; Score 43; DB 2; Length 232;
 Best Local Similarity 46.7%; Pred. No. 16;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 Oy 2 PDINPAMYXXRGIRP 16
 Db 130 PDHRAAMHLRGVLP 144

RESULT 11

T32376

hypothetical protein K10F12.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32376
 R:Wohlmann, P.; Beck, C.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid K10F12.
 A:Reference number: Z21157
 A:Accession: T32376
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-309 <WOH>
 A:Cross-references: EMBL:AF025462; PIDN:AAE1002.1; GSPDB:GN00021; CESP:K10F12.4
 A:Experimental source: strain Bristol N2; clone K10F12
 C:Genetics:
 A:Gene: CESP:K10F12.4
 A:Map position: 3
 A:Insertions: 31/3; 123/2; 196/3; 239/1

Query Match 39.4%; Score 43; DB 2; Length 309;
 Best Local Similarity 44.4%; Pred. No. 21;
 Matches 8; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
 Oy 2 PDINPAMYXXRGIRPVR 19
 Db 131 PDRSPNMYLPKS--PIGR 146

RESULT 12

A48197
 opsin, ocellar - Atlantic horseshoe crab
 C:Species: Limulus polyphemus (Atlantic horseshoe crab)
 C>Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 13-Aug-1999
 C:Accession: A48197
 R:Smith, W.C.; Price, D.A.; Greenberg, R.M.; Batteille, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6150-6154, 1993
 A:Title: Opsins from the lateral eyes and ocelli of the horseshoe crab, Limulus polyphemus
 A:Reference number: A48197; MUID:93317641; PMID:8327495
 A:Accession: A48197
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-376 <SM>
 A:Cross-references: EMBL:103792; NID:9156644; PIDD:AAA28274.1; PID:9156645
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: chromoprotein; G protein-coupled receptor; lipoprotein; photoreceptor; retin
 F:318/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 39.4%; Score 43; DB 2; Length 376;
 Best Local Similarity 46.2%; Pred. No. 26;
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 PDINPAMYXXKGI 14
 1: 11 11 11
 DB 40 PPMNPLMTSLIGV 52

RESULT 13

B48197
 opsin, lateral eye - Atlantic horseshoe crab
 C:Species: Limulus polyphemus (Atlantic horseshoe crab)
 C>Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 13-Aug-1999
 C:Accession: B48197
 R:Smith, W.C.; Price, D.A.; Greenberg, R.M.; Batteille, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6150-6154, 1993
 A:Title: Opsins from the lateral eyes and ocelli of the horseshoe crab, Limulus polyphemus
 A:Reference number: A48197; MUID:93317641; PMID:8327495
 A:Accession: B48197
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-376 <SM>
 A:Cross-references: EMBL:103791; NID:9156642; PIDD:AAA28273.1; PID:9156643
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: chromoprotein; G protein-coupled receptor; lipoprotein; photoreceptor; retin
 F:318/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 39.4%; Score 43; DB 2; Length 376;
 Best Local Similarity 46.2%; Pred. No. 26;
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 PDINPAMYXXKGI 14
 1: 11 11 11
 DB 40 PPMNPLMTSLIGV 52

RESULT 14

G64720
 probable amino acid transport protein yaaJ, sodium-dependent - Escherichia coli (strain
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: G64720
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: G64720
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-476 <BLAT>
 A:Cross-references: GB:AE000111; GB:U00096; NID:91786181; PIDD:AACT3118.1; PID:91786188;
 A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yaaJ
 C:Superfamily: sodium-dependent D-alanine/glycine transport protein
 C:Keywords: amino acid transport; transmembrane protein
 F:10-26/Domain: transmembrane #status predicted <TM>
 F:91-107/Domain: transmembrane #status predicted <TM>
 F:142-158/Domain: transmembrane #status predicted <TM>
 F:178-194/Domain: transmembrane #status predicted <TM>
 F:208-224/Domain: transmembrane #status predicted <TM>
 F:303-319/Domain: transmembrane #status predicted <TM>
 F:349-365/Domain: transmembrane #status predicted <TM>
 F:391-407/Domain: transmembrane #status predicted <TM>
 F:414-430/Domain: transmembrane #status predicted <TM>

Query Match 39.4%; Score 43; DB 2; Length 476;
 Best Local Similarity 44.4%; Pred. No. 32;
 Matches 8; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

OY 3 DIN-----PAMYXXKGI 14
 1: 11 11 11 11
 DB 120 DVNGQFRGPGPAMYXKGL 137

RESULT 15

G90629
 probable inner membrane transport protein Ecs0007 [imported] - Escherichia coli (stra
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: G90629
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
 gasawara, N.; Yasunaga, T.; Kuwahara, S.; Shiba, T.; Hattori, M.; Shinagawa, N.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: G90629
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-476 <HA>
 A:Cross-references: GB:BA000007; PIDD:BA833430.1; PID:913359463; GSPDB:GND0154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: Ecs0007
 C:Superfamily: sodium-dependent D-alanine/glycine transport protein

Query Match 39.4%; Score 43; DB 2; Length 476;
 Best Local Similarity 44.4%; Pred. No. 32;
 Matches 8; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

OY 3 DIN-----PAMYXXKGI 14
 1: 11 11 11 11
 DB 120 DVNGQFRGPGPAMYXKGL 137

Search completed: May 1, 2003, 14:31:38
 Job time: 14.375 secs

Thu May 1 15:54:22 2003

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:33:06 ; Search time 13.75 Seconds

(without alignments)
138,059 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDINPAMYXKRGIRVGRFX 22

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328255 segs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/ECT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/ECTUS_PUBCOMB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	105	96.3	20	US-09-932-161-18
2	105	96.3	31	US-09-932-161-15
3	105	96.3	87	US-10-044-592-92
4	104	95.4	20	US-09-932-161-16
5	104	95.4	20	US-09-932-161-17
6	104	95.4	20	US-10-044-592-6
7	104	95.4	20	US-10-044-592-42
8	104	95.4	21	US-10-044-592-43
9	104	95.4	22	US-10-044-592-44
10	104	95.4	31	US-09-932-161-13
11	104	95.4	31	US-09-932-161-14
12	104	95.4	31	US-10-044-592-4
13	104	95.4	31	US-10-044-592-5
14	104	95.4	31	US-10-044-592-39
15	104	95.4	32	US-10-044-592-40
16	104	95.4	33	US-10-044-592-41
17	104	95.4	70	US-10-044-592-80
18	104	95.4	82	US-10-044-592-1
19	104	95.4	86	US-10-044-592-96

20	104	95.4	91	US-10-044-592-94	Sequence 94, Appl
21	104	95.4	98	US-10-044-592-28	Sequence 28, Appl
22	104	95.4	98	US-10-044-592-38	Sequence 38, Appl
23	104	95.4	98	US-10-044-592-82	Sequence 82, Appl
24	104	95.4	98	US-10-044-592-84	Sequence 84, Appl
25	104	95.4	98	US-10-044-592-86	Sequence 86, Appl
26	104	95.4	98	US-10-044-592-88	Sequence 88, Appl
27	98	89.9	19	US-10-044-592-27	Sequence 27, Appl
28	93	85.3	29	US-10-044-592-26	Sequence 26, Appl
29	83	76.1	25	US-10-044-592-78	Sequence 78, Appl
30	57	52.3	9	US-10-044-592-8	Sequence 8, Appl
31	48	44.0	10	US-10-044-592-9	Sequence 9, Appl
32	45	41.3	209	US-10-108-915-30	Sequence 30, Appl
33	44	40.4	276	US-09-925-302-535	Sequence 535, App
34	43	39.4	476	US-09-815-242-10014	Sequence 10014, A
35	42	38.5	269	US-09-738-636-4086	Sequence 4086, Ap
36	42	38.5	604	US-09-738-626-3886	Sequence 3886, Ap
37	41	37.6	236	US-09-934-392-4	Sequence 4, Appl
38	41	37.6	417	US-09-738-626-4642	Sequence 4642, Ap
39	41	37.6	485	US-09-815-242-13712	Sequence 13712, A
40	41	37.6	753	US-09-888-615-68	Sequence 68, Appl
41	40	36.7	89	US-08-764-877-1227	Sequence 1227, Ap
42	40	36.7	468	US-09-815-242-5144	Sequence 5144, Ap
43	39.5	36.2	116	US-09-939-825-26	Sequence 26, Appl
44	39.5	36.2	415	US-09-924-256A-26	Sequence 26, Appl
45	39	35.8	40	US-10-044-592-80	Sequence 80, Appl

ALIGNMENTS

RESULT 1

US-09-932-161-18

Sequence 18, Appl Application US/09932161

Patent No. US20020037533A1

GENERAL INFORMATION:

APPLICANT: Clivell, Olivier

APPLICANT: Lin, Steven

TITLE OF INVENTION: Screening and Therapeutic Methods For

TITLE OF INVENTION: Promoting Wakefulness and Sleep

FILE REFERENCE: P-UC 4679

CURRENT APPLICATION NUMBER: US/09/932,161

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: US 09/560,915

PRIOR FILING DATE: 2000-04-28

NUMBER OF SEQ ID NOS: 24

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 18

TYPE: PRT

LENGTH: 20

ORGANISM: Homo Sapien

US-09-932-161-18

Query Match

Best Local Similarity 96.3%, Score 105, DB 10, Length 20,

Matches 18, Conservative 0, Pred. No. 5.2e-10,

Mismatches 2, Indels 0, Gaps 0,

OY 1 TPDINPAMYXKRGIRVGRFX 20

DB 1 TPDINPAMYXKRGIRVGRFX 20

RESULT 2

US-09-932-161-15

Sequence 15, Appl Application US/09932161

Patent No. US20020037533A1

GENERAL INFORMATION:

APPLICANT: Clivell, Olivier

APPLICANT: Lin, Steven

TITLE OF INVENTION: Screening and Therapeutic Methods For

TITLE OF INVENTION: Promoting Wakefulness and Sleep

FILE REFERENCE: P-UC 4679

CURRENT APPLICATION NUMBER: US/09/932,161

;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 15
;; LENGTH: 31
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-932-161-15

Query Match 96.3%; Score 105; DB 10; Length 31;
Best Local Similarity 90.0%; Pred. No. 9e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 12 TPDINPAMYASRGIRPVGRF 31

RESULT 3

US-10-044-592-92
;; Sequence 92, Application US/10044592
;; Patent No. US20020143152A1
;; GENERAL INFORMATION:
;; APPLICANT: Hinuma, Shuji
;; TITLE OF INVENTION: Polypeptides, their Production and Use
;; FILE REFERENCE: 2463US2P
;; CURRENT APPLICATION NUMBER: US/10/044,592
;; CURRENT FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 09/403639
;; PRIOR FILING DATE: 1999-25-10
;; PRIOR APPLICATION NUMBER: PCT/JP98/01923
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: JP 9-109974
;; PRIOR FILING DATE: 1997-04-28
;; NUMBER OF SEQ ID NOS: 96
;; SOFTWARE:
;; SEQ ID NO 92
;; LENGTH: 87
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 96.3%; Score 105; DB 12; Length 87;
Best Local Similarity 90.0%; Pred. No. 2.2e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 34 TPDINPAMYASRGIRPVGRF 53

RESULT 4

US-09-932-161-16
;; Sequence 16, Application US/09932161
;; Patent No. US20020037533A1
;; GENERAL INFORMATION:
;; APPLICANT: Civeilli, Olivier
;; APPLICANT: Lin, Steven
;; TITLE OF INVENTION: Screening and Therapeutic Methods For
;; FILE REFERENCE: P-UC 4679
;; CURRENT APPLICATION NUMBER: US/09/932,161
;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 16
;; LENGTH: 20
;; TYPE: PRT

;; ORGANISM: Bos taurus
US-09-932-161-16

Query Match 95.4%; Score 104; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 5

US-09-932-161-17
;; Sequence 17, Application US/09932161
;; Patent No. US20020037533A1
;; GENERAL INFORMATION:
;; APPLICANT: Civeilli, Olivier
;; APPLICANT: Lin, Steven
;; TITLE OF INVENTION: Screening and Therapeutic Methods For
;; FILE REFERENCE: P-UC 4679
;; CURRENT APPLICATION NUMBER: US/09/932,161
;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 17
;; LENGTH: 20
;; TYPE: PRT
;; ORGANISM: Rattus
US-09-932-161-17

Query Match 95.4%; Score 104; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 6

US-10-044-592-6
;; Sequence 6, Application US/10044592
;; Patent No. US20020143152A1
;; GENERAL INFORMATION:
;; APPLICANT: Hinuma, Shuji
;; APPLICANT: Fukusumi, Shoji
;; TITLE OF INVENTION: Polypeptides, their Production and Use
;; FILE REFERENCE: 2463US2P
;; CURRENT APPLICATION NUMBER: US/10/044,592
;; CURRENT FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 09/403639
;; PRIOR FILING DATE: 1999-25-10
;; PRIOR APPLICATION NUMBER: PCT/JP98/01923
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: JP 9-109974
;; PRIOR FILING DATE: 1997-04-28
;; NUMBER OF SEQ ID NOS: 96
;; SOFTWARE:
;; SEQ ID NO 6
;; LENGTH: 20
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; NAME/KEY: PEPTIDE
;; LOCATION: (1)..(20)
;; OTHER INFORMATION: antigen
US-10-044-592-6

Query Match 95.4%; Score 104; DB 12; Length 20;

Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 1 TPDINPAMYTGIRGIRPVGRF 20

RESULT 7
US-10-044-592-42
; Sequence 42, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Minuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 42
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-42

Query Match 95.4%; Score 104; DB 12; Length 20;
Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 8
US-10-044-592-43
; Sequence 43, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Minuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 43
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-43

Query Match 95.4%; Score 104; DB 12; Length 21;
Best Local Similarity 90.0%; Pred. No. 7.8e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||

Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 9
US-10-044-592-44
; Sequence 44, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Minuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 44
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-44

Query Match 95.4%; Score 104; DB 12; Length 22;
Best Local Similarity 90.0%; Pred. No. 8.1e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 10
US-09-932-161-13
; Sequence 13, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:
; APPLICANT: Civeill, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods for
; FILE REFERENCE: P-UC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-932-161-13

Query Match 95.4%; Score 104; DB 10; Length 31;
Best Local Similarity 90.0%; Pred. No. 1.1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 12 TPDINPAMYAGRGIRPVGRF 31

RESULT 11
US-09-932-161-14
; Sequence 14, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:

APPLICANT: Civeilli, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods For
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 31
TYPE: PRT
ORGANISM: Rattus
US-09-932-161-14

Query Match 95.4%; Score 104; DB 10; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMVTGRIIPVGRF 31

RESULT 12
US-10-044-592-4
Sequence 4, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Murine
US-10-044-592-4

Query Match 95.4%; Score 104; DB 12; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMVTGRIIPVGRF 31

RESULT 13
US-10-044-592-5
Sequence 5, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 5
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(31)
OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 95.4%; Score 104; DB 12; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMVTGRIIPVGRF 31

RESULT 14
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 95.4%; Score 104; DB 12; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMVTGRIIPVGRF 31

RESULT 15
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923

Thu May 1 15:54:22 2003

us-09-446-543a-73.rapb

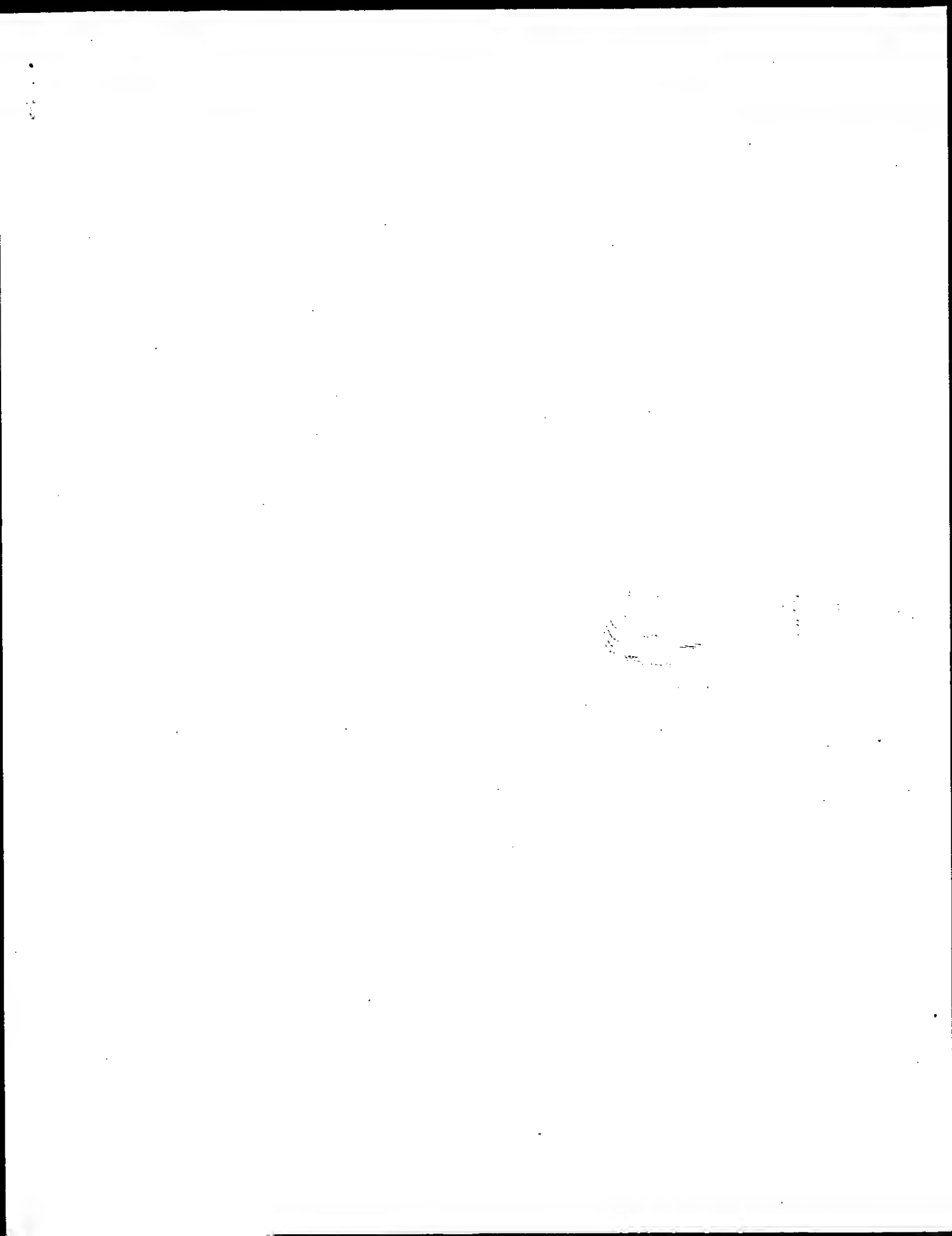
Page 5

; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 40
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-40

Query Match 95.4%; Score 104; DB 12; Length 32;
Best Local Similarity 90.08; Pred. No. 1.2e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPYGRF 20
||||||| |||||||
Db 12 TPDINPAMYAGRGIRPYGRF 31

Search completed: May 1, 2003, 14:46:06
Job time : 13.75 secs



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:13 ; Search time 10.6562 Seconds

(Without alignments)
60,744 Million cell updates/sec

Title: US-09-446-543a-73

Sequence: 1 TPDINPMWXXRGIRVGRFX 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	20	3	US-09-105-678A-46
2	105	96.3	20	4	US-08-776-971-64
3	105	96.3	20	4	US-09-421-208-46
4	105	96.3	20	4	US-09-560-915-18
5	105	96.3	21	3	US-09-105-678A-47
6	105	96.3	21	4	US-08-776-971-65
7	105	96.3	21	4	US-09-421-208-47
8	105	96.3	22	3	US-09-105-678A-48
9	105	96.3	22	4	US-08-776-971-66
10	105	96.3	22	4	US-09-421-208-48
11	105	96.3	31	3	US-09-105-678A-9
12	105	96.3	31	3	US-09-105-678A-43
13	105	96.3	31	4	US-08-776-971-61
14	105	96.3	31	4	US-09-421-208-9
15	105	96.3	31	4	US-09-421-208-43
16	105	96.3	31	4	US-09-560-915-15
17	105	96.3	32	3	US-09-105-678A-44
18	105	96.3	32	4	US-08-776-971-62
19	105	96.3	32	4	US-09-421-208-44
20	105	96.3	33	3	US-09-105-678A-45
21	105	96.3	33	4	US-08-776-971-63
22	105	96.3	33	4	US-09-421-208-45
23	105	96.3	87	4	US-08-776-971-59
24	105	96.3	87	4	US-08-776-971-135
25	105	96.3	87	4	US-08-776-971-138
26	104	95.4	20	3	US-09-105-678A-34
27	104	95.4	20	3	US-09-105-678A-40

28	104	95.4	20	4	US-08-776-971-8	Sequence 8, Appl
29	104	95.4	20	4	US-08-776-971-50	Sequence 50, Appl
30	104	95.4	20	4	US-08-776-971-98	Sequence 98, Appl
31	104	95.4	20	4	US-09-421-208-34	Sequence 34, Appl
32	104	95.4	20	4	US-09-421-208-40	Sequence 40, Appl
33	104	95.4	20	4	US-09-560-915-16	Sequence 16, Appl
34	104	95.4	20	4	US-09-560-915-17	Sequence 17, Appl
35	104	95.4	21	3	US-09-105-678A-35	Sequence 35, Appl
36	104	95.4	21	3	US-08-776-971-9	Sequence 9, Appl
37	104	95.4	21	4	US-08-776-971-51	Sequence 51, Appl
38	104	95.4	21	4	US-09-421-208-35	Sequence 35, Appl
39	104	95.4	21	4	US-09-421-208-41	Sequence 41, Appl
40	104	95.4	22	3	US-09-105-678A-36	Sequence 36, Appl
41	104	95.4	22	3	US-09-105-678A-42	Sequence 42, Appl
42	104	95.4	22	4	US-08-776-971-10	Sequence 10, Appl
43	104	95.4	22	4	US-08-776-971-52	Sequence 52, Appl
44	104	95.4	22	4	US-09-421-208-36	Sequence 36, Appl
45	104	95.4	22	4	US-09-421-208-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-105-678A-46
Sequence 46, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takao
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-46
Query Match 96.3%, Score 105, DB 3, Length 20,
Best Local Similarity 90.0%, Pred. No. 2.9e-11,
Matches 18, Conservative 0, Mismatches 2, Indels 0, Gaps 0,
QY 1 TPDINPMWXXRGIRVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 2

US-08-776-971-64

Sequence 64, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: NINUMA, Shuji

Nabata, Yogo

Kawamata, Yuji

Hosoya, Masaki

Fuji, Ryo

Fukusumi, Shoji

Kitada, Chieko

TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 64:

US-08-776-971-64

Query Match 96.3%; Score 105; DB 4; Length 20;

Best Local Similarity 90.0%; Pred. No. 2,9e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYASRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 3

US-09-421-208-46

Sequence 46, Application US/09421208

Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Noriya, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/421,208

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/105,678

FILING DATE: 26-JUN-1998

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-421-208-46

Query Match 96.3%; Score 105; DB 4; Length 20;

Best Local Similarity 90.0%; Pred. No. 2,9e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYASRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 4

US-09-560-915-18

Sequence 18, Application US/09560915

Patent No. 6383764

GENERAL INFORMATION:

APPLICANT: Clivell, Olivier

APPLICANT: Lin, Steven

TITLE OF INVENTION: Therapeutic Compositions and Methods

FILE REFERENCE: P-00 3534

CURRENT APPLICATION NUMBER: US/09/560,915

NUMBER OF SEQ ID NOS: 24

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 18

LENGTH: 20

TYPE: PRT

ORGANISM: Homo sapien

US-09-560-915-18

Query Match 96.3%; Score 105; DB 4; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.9e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXKRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 5

US-09-105-678A-47
Sequence 47, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-47

Query Match 96.3%; Score 105; DB 3; Length 21;
Best Local Similarity 90.0%; Pred. No. 3.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXKRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 6

US-08-776-971-65
Sequence 65, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Nishimura, Osamu

APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

Nosoya, Masaki

Fuji, Ryo

Fukushima, Shoji

Kikada, Chieko

TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 65:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 65:

US-08-776-971-65

Query Match 96.3%; Score 105; DB 4; Length 21;
Best Local Similarity 90.0%; Pred. No. 3.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXKRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 7

US-09-421-208-47
Sequence 47, Application US/09421208
Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Tanaka, Yoko

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-47

Query Match 96.3%; Score 105; DB 4; Length 21;
Best Local Similarity 90.0%; Fred. No. 3.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPMATYXRGIRPVGRF 20
DB 1 TPDINPMATYXRGIRPVGRF 20

RESULT 8
US-09-105-678A-48
Sequence 48, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-48

Query Match 96.3%; Score 105; DB 3; Length 22;
Best Local Similarity 90.0%; Fred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPMATYXRGIRPVGRF 20
DB 1 TPDINPMATYXRGIRPVGRF 20

RESULT 9
US-08-776-971-66
Sequence 66, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hama, Shuji
Hakata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Ekusumi, Shoji
Kikada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 66:

SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 66
US-08-776-971-66

Query Match 96.3%; Score 105; DB 4; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMWYXXRGIRPVGRF 20
||||| |||||
Db 1 TPDINPAMWYXXRGIRPVGRF 20

RESULT 10

US-09-421-208-48
Sequence 48, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-48

Query Match 96.3%; Score 105; DB 4; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMWYXXRGIRPVGRF 20
||||| |||||
Db 1 TPDINPAMWYXXRGIRPVGRF 20

RESULT 11
US-09-105-678A-9
Sequence 9, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-9

Query Match 96.3%; Score 105; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMWYXXRGIRPVGRF 20
||||| |||||
Db 12 TPDINPAMWYXXRGIRPVGRF 31

RESULT 12
US-09-105-678A-43
Sequence 43, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-43

Query Match 96.3%; Score 105; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 4,8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXRGIRPVGRF 20
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 13
US-08-776-971-61
Sequence 61, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Hatake, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-08-776-971-61

Query Match 96.3%; Score 105; DB 4; Length 31;
Best Local Similarity 90.0%; Pred. No. 4,8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXRGIRPVGRF 20
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 14
US-09-421-208-9
Sequence 9, Application US/09421208
Patent No. 6258361
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
Moriya, Takeo
Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-9

Query Match 96.3%; Score 105; DB 4; Length 31,
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
DB 12 TPDINPAMYASRGIRPVGRF 31

RESULT 15
US-09-421-208-43
Sequence 43, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coilln, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-43

Query Match 96.3%; Score 105; DB 4; Length 31,
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
DB 12 TPDINPAMYASRGIRPVGRF 31

Search completed: May 1, 2003, 14:33:54
Job time: 10.6562 secs

10

GenCore version 5.1.4-578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 1, 2003, 14:29:12 ; Search time 33.9062 Seconds

(without alignments)
188.386 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171
Sequence: 1 SRRHSHMEIRTPDINPAMYASRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158	92.4	98	6 Q8WN12	Q8WN12 ovine aries
2	98	57.3	117	13 Q9W624	Q9W624 carassius a
3	56.5	33.0	664	13 Q911W4	Q911W4 pseudomonas
4	55	32.2	315	2 O87474	O87474 burkholderi
5	54	31.6	790	10 Q9H371	Q9H371 arabidopsis
6	53.5	31.3	314	16 O06348	O06348 mycobacteri
7	53	31.0	327	10 Q94K99	Q94K99 brassica ca
8	53	31.0	327	10 Q94F29	Q94F29 caulobacter
9	52	30.4	785	17 Q9A5E9	Q9A5E9 methanobact
10	51.5	30.1	420	2 Q931Z7	Q931Z7 streptomyce
11	51	29.8	54	4 Q9UJF9	Q9UJF9 homo sapien
12	51	29.8	277	11 Q9DBA2	Q9DBA2 mus musculu
13	51	29.8	286	11 Q9DAB5	Q9DAB5 mus musculu
14	51	29.8	450	11 Q9DA19	Q9DA19 mus musculu
15	51	29.8	450	11 Q9DA19	Q9DA19 mus musculu
16	51	29.8	465	4 O60687	O60687 homo sapien

17	51	29.8	465	4 Q8WN85	Q8WN85 homo sapien
18	50	29.2	123	2 Q45883	Q45883 eubacterium
19	50	29.2	128	16 P74747	P74747 synecocyst
20	50	29.2	692	2 Q918U6	Q918U6 rhodospirill
21	50	29.2	939	5 Q9GRA0	Q9GRA0 hemocentrot
22	50	29.2	1253	2 Q8RTV7	Q8RTV7 uncultured
23	49.5	28.9	176	3 Q08689	Q08689 saccharomyc
24	49	28.7	173	16 Q9EYB3	Q9EYB3 escherichia
25	49	28.7	173	17 Q97A00	Q97A00 thermoplasma
26	49	28.7	326	2 Q9F642	Q9F642 stigmatella
27	48.5	28.4	303	16 Q8YWC7	Q8YWC7 anabaena sp
28	48.5	28.4	1501	10 Q9SD86	Q9SD86 arabidopsis
29	48	28.1	118	16 Q8UKM3	Q8UKM3 agrobacteri
30	48	28.1	220	16 Q91022	Q91022 pseudomonas
31	48	28.1	335	2 Q88002	Q88002 bordetella
32	48	28.1	335	2 Q45375	Q45375 bordetella
33	48	28.1	348	5 Q45431	Q45431 caenorhabdit
34	48	28.1	455	16 Q33340	Q33340 drosophila
35	48	28.1	514	5 Q961A3	Q961A3 drosophila
36	48	28.1	596	5 Q9VP62	Q9VP62 drosophila
37	48	28.1	863	5 Q9VRV3	Q9VRV3 drosophila
38	48	28.1	1084	5 Q9TYW4	Q9TYW4 caenorhabdit
39	47.5	27.8	184	11 Q8YH88	Q8YH88 mus musculu
40	47.5	27.8	273	11 Q8R5B6	Q8R5B6 mus musculu
41	47.5	27.8	333	16 Q9RUI0	Q9RUI0 streptomyce
42	47.5	27.8	401	16 Q8UHR0	Q8UHR0 agrobacteri
43	47.5	27.8	503	16 Q9KRY1	Q9KRY1 vibrio chol
44	47.5	27.8	805	5 Q9T253	Q9T253 drosophila
45	47	27.5	145	5 Q95ZB5	Q95ZB5 leishmania

ALIGNMENTS

RESULT 1
ID Q8WN12 PRELIMINARY: PRT: 98 AA.
AC Q8WN12;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Preprolactin-releasing peptide.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Curlew J.D., Rusters D.H.L., Barclay J.L., Anderson S.T.;
RT "Prolactin-releasing peptide (PRP) in the ewe: cDNA cloning, mRNA
RT distribution, and effects on prolactin secretion in vitro and in
RT vivo."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF450453; AAL47178.1; ZAS3331ED62CAAB5 CRC64;
SQ SEQUENCE 98 AA: 10513 MW: ZAS3331ED62CAAB5 CRC64;

Query Match 92.4%; Score 158; DB 6; Length 98;
Best Local Similarity 90.3%; Pred. No. 1.2e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRRHSHMEIRTPDINPAMYASRGIRPVGRF 31
DB 23 SRRHSHMEIRTPDINPAMYASRGIRPVGRF 53

RESULT 2
ID Q9W624 PRELIMINARY: PRT: 117 AA.
AC Q9W624;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE C-RF amide.
 OS Carassius auratus (Goldfish).
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Carassius.
 RN NCBI_TaxID=7957;
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Satake H., Minakata H., Fujimoto M.;
 RT "Carassius kramidi (C-RF amide).";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB020024; BAA76662.1; -
 SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;
 Query Match 57.3%; Score 98; DB 13; Length 117;
 Best Local Similarity 53.0%; Pred. No. 9.4e-07;
 Matches 14; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 QY 6 HSMETRPDINPAWYASRGIRPGRF 31
 DB 50 HNDNRSPEIDPTVYGRVPIGRF 75
 RESULT 3
 Q9JIM4 PRELIMINARY; PRT; 664 AA.
 AC 09JIM4;
 DT 01-MAR-2001 (TRENBLREL. 16, Created)
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DE 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 GN PA2151.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 RX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mitsuuchi S.D., Warriner P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuen Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 RL EMBL: AE004642; AG05539.1;
 DR InterPro: IPR001589; Actbind_actin.
 DR InterPro: IPR000461; Alpha_amyase.
 DR Pfam: PF00128; alpha-amyase; 1.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 664 AA; 76329 MW; 8F59FEED54C308AD CRC64;
 Query Match 33.0%; Score 56.5; DB 16; Length 664;
 Best Local Similarity 45.8%; Pred. No. 7.6;
 Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;
 QY 4 HRHSNETPDINPAWYASRGIRP 27
 DB 470 YRPNFVWTPDINP-WFLRSGRP 492
 RESULT 4
 ID 087474 PRELIMINARY; PRT; 315 AA.
 AC 087474;
 DT 01-NOV-1998 (TRENBLREL. 08, Created)
 DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)

DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Trihydroxytoluene oxygenase.
 GN DNDP.
 OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; Beta subdivision; Burkholderia group;
 OC Burkholderia.
 RN NCBI_TaxID=292;
 RP SEQUENCE FROM N.A.
 RC STRAIN=DNT;
 RA MEDLINE=93194809; PubMed=8449889;
 RA Suen W.C., Spain J.C.;
 RT "Cloning and characterization of Pseudomonas sp. strain DNT genes for
 2,4-dinitrotoluene degradation.";
 RL J. Bacteriol. 175:1831-1837(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DNT;
 RX MEDLINE=99121037;
 RA Halgler B.E., Johnson G.R., Suen W.C., Spain J.C.;
 RT "Biochemical and genetic evidence for meta-ring cleavage of 2,4,5-
 trihydroxytoluene in Burkholderia sp. strain DNT.";
 RL J. Bacteriol. 181:965-972(1999).
 DR EMBL: AF076848; AAD12738.1; -
 SQ SEQUENCE 315 AA; 34745 MW; E99261179022961E CRC64;
 Query Match 32.2%; Score 55; DB 2; Length 315;
 Best Local Similarity 44.4%; Pred. No. 5.7;
 Matches 16; Conservative 3; Mismatches 11; Indels 6; Gaps 2;
 QY 1 STHHSNETPDP--INPAWYASRGIRP--VGR 30
 DB 216 SRHHSCFEIDIDDAQFLGNKWLASRGKPGWGR 251
 RESULT 5
 Q9N371 PRELIMINARY; PRT; 790 AA.
 AC 09N371;
 DT 01-OCT-2000 (TRENBLREL. 15, Created)
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
 DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
 GN F15616.60.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 RX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA De Haan M., Maarse A.C., Grivell L.A., Meves H.W., Lemcke K.,
 Mayer K.F.X., Queller F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132959; CAB71097.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 790 AA; 87376 MW; B222724B75630F30 CRC64;
 Query Match 31.6%; Score 54; DB 10; Length 790;
 Best Local Similarity 52.6%; Pred. No. 21;
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 13 PDINPAWYASRGIRPGRF 31
 DB 366 PPHNPRTYSGRIQPHGRW 384
 RESULT 6
 ID 006348

```

ID 006348; PRELIMINARY; PRT: 314 AA.
AC 006348;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 33.2 kDa protein (Oxidoreductase, short-chain
DE dehydrogenase/reductase family).
GN RV3485C OR MT3589 OR MTCYL3E12.38C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Elismeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RS SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / OSBKOSH;
RC Fietischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Debay R., Dodson R., Gwinn M., Hart D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC EMBL: Z95390; CAB08708.1; -.
DR DR EMBL: AE007162; AAK47948.1; ALT_INIT.
DR HSSP: P29132; IDPI.
DR TRIR: MT3589; -.
DR Tuberculin; RV3485C; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN.1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 314 Aa; 33194 MW; 26144BA917E09274 CRC64;
OY 1 SRTSR-----HSNEIRPDINPANTASRGIRP 27
DY 1 111 1 1::: 111
DB 187 SWTHRFAGAYGTSAVDHMKRLADELGPVWVRSIRP 226
Matches 11; Conservative 5; Mismatches 11; Indels 13; Gaps 1
OY 1 SRTSR-----HSNEIRPDINPANTASRGIRP 27
DY 1 111 1 1::: 111
DB 187 SWTHRFAGAYGTSAVDHMKRLADELGPVWVRSIRP 226
Query Match 31.3%; Score 53.5; DB 16; Length 314;
Best Local Similarity 27.5%; Pred. No. 9.4;
Matches 11; Conservative 5; Mismatches 11; Indels 13; Gaps 1
OY 1 SRTSR-----HSNEIRPDINPANTASRGIRP 27
DY 1 111 1 1::: 111
DB 187 SWTHRFAGAYGTSAVDHMKRLADELGPVWVRSIRP 226
RESULT 7
ID 094KU9 PRELIMINARY; PRT: 327 AA.
AC 094KU9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PlastiD-lipid associated protein PAP1.
DE Brassica campestris (Field mustard).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxId=3711;

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-21249173; PubMed=11351096;
RA Kim H.U., Wu S.S., Ratnayake C., Huang A.H.;
RT "Brassica rapa Has Three Genes That Encode Proteins Associated with
  Different Neutral Lipids in Plastids of Specific Tissues.";
RL Plant Physiol. 126:330-341(2001).
DR EMBL: AF290563; AAK57561.1; -.
SQ SEQUENCE 327 AA; 35678 MW; 8C7B87FCD6C02422 CRC64;

OY 4 HRHSMKRTPDINPAW 19
   |||:::|:-|:-|
Db 45 HRHDFKVRASDVNDW 60

RESULT 8
O94FE9 ID O94FE9 PRELIMINARY; PRT; 327 AA.
AC O94FE9;
DT 01-DEC-2001 (T:EMBLrel. 19, Created)
DT 01-DEC-2001 (T:EMBLrel. 19, last sequence update)
DE 01-DEC-2001 (T:EMBLrel. 19, last annotation update)
DE Plastid-lipid associated protein PAP1.
OC Brassica campestris (Field mustard).
OC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eumetids II; Brassicales; Brassicaceae; Brassica.
CX NCBI_TaxID=3711;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE-21249173; PubMed=11351096;
RA Kim H.U., Wu S.S., Ratnayake C., Huang A.H.;
RT "Brassica rapa Has Three Genes That Encode Proteins Associated with
  Different Neutral Lipids in Plastids of Specific Tissues.";
RL Plant Physiol. 126:330-341(2001).
DR EMBL: AF290566; AAK57564.1; -.
SQ SEQUENCE 327 AA; 35644 MW; 6116E7F1B6C02C88 CRC64;

Query Match 31.0%; Score 53; DB 10; Length 327;
Best Local Similarity 43.8%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0

OY 4 HRHSMKRTPDINPAW 19
   |||:::|:-|:-|
Db 45 HRHDFKVRASDVNDW 60

RESULT 9
O9A5E9 ID O9A5E9 PRELIMINARY; PRT; 637 AA.
AC O9A5E9;
DT 01-JUN-2001 (T:EMBLrel. 17, Created)
DT 01-JUN-2001 (T:EMBLrel. 17, last sequence update)
DT 01-JUN-2002 (T:EMBLrel. 21, last annotation update)
DE sensory box histidine kinase/response regulator.
GN CC3501.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE-21173696; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocky J., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Helt D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Knouri H., Shetty J., Berry K.,

```

RA Uterback T., Tran K., Wolf A., Vamathavan J., Ertolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RT "Complete genome sequence of *Caulobacter crescentus*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
 CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
 CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
 DR EMBL: AE005918; AAK24472.1;
 DR HSSP: O56312; 3TWY.
 DR TIGR: CC2501;
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; hact_sens_pr_c.
 DR InterPro: IPR003661; HIS_KIN_519.
 DR InterPro: IPR004359; HIS_KIN_519.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000700; PAS-assoc_C.
 DR InterPro: IPR000014; PAS_domain.
 DR InterPro: IPR001789; Response_reg.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00785; PAC; 1.
 DR Pfam: PF00989; PAS; 1.
 DR Pfam: PF00072; response_reg; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR PRODOM: PD000039; Response_reg; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00388; B1AKA; 1.
 DR SMART: SM00086; PAC; 1.
 DR SMART: SM00091; PAS; 1.
 DR SMART: SM00448; REC; 1.
 DR TIGRFAMS: TIGR00229; sensory_box; 1.
 DR kinase: Phosphorylation; sensory transduction; Transferase;
 KM Complete proteome.
 SO SEQUENCE 637 AA; 68511 MW; 0EDEBAF76FPA8611 CRC64;
 Query Match 31.0%; Score 53; DB 16; Length 637;
 Best Local Similarity 48.0%; Pred. No. 24;
 Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
 QY 4 HRHSMETRPDINPAWYASRGIRPV 28
 DB 22 HRSDSLRSPALNPALRVLRAV 46
 |||:::|||||1111
 RESULT 10
 ID 026276 PRELIMINARY; PRT; 785 AA.
 AC 026276;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sensory transduction histidine kinase.
 GN MTN17.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID-187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE-98037514; PubMed-9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-W., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Yicare L., Keagle P., Lumm W., Polther B., Olu D.,
 RA Spadefora R., Viscare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-T., Rice P., Noelling J., Reeve J.N.,
 RT "Complete genome sequence of *Methanobacterium thermoautotrophicum*
 RT deltaH: functional analysis and comparative genomics,"
 RL J. Bacteriol. 179:7133-7155(1997).
 DR EMBL: AE000805; AAB84680.1;
 DR InterPro: IPR003594; ATPbind_ATPase.

DR InterPro: IPR000515; BPD_transp.
 DR InterPro: IPR004359; HIS_KIN_519.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000700; PAS-assoc_C.
 DR InterPro: IPR000014; PAS_domain.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00785; PAC; 1.
 DR Pfam: PF00989; PAS; 1.
 DR Pfam: PF00072; response_reg; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00086; PAC; 1.
 DR SMART: SM00091; PAS; 1.
 DR TIGRFAMS: TIGR00229; sensory_box; 3.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 SW Complete proteome.
 SO SEQUENCE 785 AA; 87726 MW; ADD502C928307986 CRC64;
 Query Match 30.4%; Score 52; DB 17; Length 785;
 Best Local Similarity 46.4%; Pred. No. 42;
 Matches 13; Conservative 1; Mismatches 12; Indels 2; Gaps 1;
 QY 4 HRHSMETRPDINPAWYASRGIRPV 29
 DB 412 HRATPRIRPDGNYRWVEYDRIRTDG 439
 |||:::|||||1111
 RESULT 11
 ID 093127 PRELIMINARY; PRT; 420 AA.
 AC 093127;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Chain length factor-like protein.
 GN AUR28.
 OS Streptomyces aureofaciens.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID-1894;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CCM3239.
 RA Kormanec J., Bistakova J., Novakova R., Homeroova D., Rezuchova B.,
 RT "Cloning and characterization of a new polyketide gene cluster in
 RT streptomyces aureofaciens CCM3239,"
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY033994; AAK61719.1;
 DR InterPro: IPR000794; Ketoacyl-synt.
 DR Pfam: PF00109; ketoacyl-synt; 1.
 DR Pfam: PF02801; ketoacyl-synt_C; 1.
 SO SEQUENCE 420 AA; 43011 MW; 3C27E22BE86C2DEA CRC64;
 Query Match 30.1%; Score 51.5; DB 2; Length 420;
 Best Local Similarity 57.9%; Pred. No. 25;
 Matches 11; Conservative 2; Mismatches 1; Indels 5; Gaps 1;
 QY 18 AWYAS----RGIRPVGRP 31
 DB 37 AWYAAVLRGSGIRPVGRP 55
 |||:::|||||1111
 RESULT 12
 ID 090JF9 PRELIMINARY; PRT; 54 AA.
 AC 090JF9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE DJ479J7.3 (Sushi-repeat protein (SRPUL)) (Fragment).
 GN DJ479J7.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homalidae; Homo.
 OX NCBI_TaxID-9606;

DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE 1700023B02R1K protein.
 GN 1700023B02R1K.
 DS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RC MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh N., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
 RA Saito T., Dkazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Dkido T., Furuno M., Aono H., Baldarelli R., Barns G.,
 RA Blake J., Boffelli D., Bojunga N., Garnciel P., de Bonaldo M.F.,
 RA Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,
 RA Lyons P., Marchionni L., Mashima J., Hume D.A., Kamuya M., Lee N.H.,
 RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakanoto N.,
 RA Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki B., Toyooka K., Wang K.B., Waltz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL EMBL; AK005900; BAB24307.1; -
 DR MGP; MGI:1914185; 1700023B02R1K.
 DR PRINTS; PR01574; TUBBYPROTEIN.
 SO SEQUENCE 286 AA; 32818 MW; 068003C5E894827B CRC64;
 Duerly Match 29.8%; Score 51; DB 11; Length 286;
 Best Local Similarity 46.7%; Pred. No. 20;
 Matches 14; Conservative 2; Mismatches 8; Indels 6; Gaps 1.
 Cy 1 SRTHSHMEIIRPPDINPMVTSKRGIRPMP 30
 Db 209 SRSRHSPEKSGSDRN-----RGIRSR 232
 RESULT 15
 C9DA19 PRELIMINARY; PRT; 450 AA.
 AC C9DA19;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 GN 1700023B02R1K protein.
 GN 1700023B02R1K.
 OS Mus musculus (Mouse).
 DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 DC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 DX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RC MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh N., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
 RA Saito T., Dkazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barns G.,
 RA Blake J., Boffelli D., Bojunga N., Garnciel P., de Bonaldo M.F.,
 RA Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,
 RA Lyons P., Marchionni L., Mashima J., Hume D.A., Kamuya M., Lee N.H.,
 RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakanoto N.,
 RA Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki B., Toyooka K., Wang K.B., Waltz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL EMBL; AK005900; BAB24307.1; -
 DR MGP; MGI:1914185; 1700023B02R1K.
 DR PRINTS; PR01574; TUBBYPROTEIN.
 SO SEQUENCE 286 AA; 32818 MW; 068003C5E894827B CRC64;

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustjuncich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokura K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK006260; BAB24488.1;
 DR MGI; MGI:1914185; 1700023B02R1.
 DR InterPro; IPR000345; Cytc_heme_bind.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNORM.1.
 SO SEQUENCE 450 AA; 5185 MW; F32F118ED6DA4EAC CRC64;

Query Match 29.8%; Score 51; DB 11; Length 450;
 Best Local Similarity 46.7%; Pred. No. 32;
 Matches 14; Conservative 2; Mismatches 8; Indels 6; Gaps 1;

QY 1 SETHHSNEIRTPDINPAMYASRGIRPYGR 30
 Db 373 SRSHRSEPKSGSDRN-----RGIHSRSR 396

Search completed: May 1, 2003, 14:32:53
 Job time : 38.9062 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12, Search time 17.4375 Seconds
(without alignments)
170.906 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171
Sequence: 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1: PIR73:*
2: PIR1:*
3: PIR2:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	87.1	83	JC7607	prolactin-releasing
2	56.5	33.0	664	F83376	conserved hypothet
3	54	31.6	790	T47959	hypothetical prote
4	53.5	31.3	314	B70569	hypothetical prote
5	53	31.0	637	D87558	sensory box histid
6	52	30.4	785	F69099	sensory transducti
7	50.5	29.5	1882	S73484	hypothetical prote
8	50	29.2	123	S77900	hypothetical prote
9	50	29.2	128	S76955	hypothetical prote
10	49.5	28.9	176	S67150	hypothetical prote
11	49	28.7	72	E91002	hypothetical prote
12	48.5	28.4	303	AH2016	probable regulator
13	48.5	28.4	1501	T45623	hypothetical prote
14	48	28.1	118	AC3169	hypothetical prote
15	48	28.1	220	C83292	hypothetical prote
16	48	28.1	335	S70671	probable glutathio
17	48	28.1	348	T21648	lipopolysaccharide
18	48	28.1	455	D70885	hypothetical prote
19	48	28.1	1084	T33759	probable aldol prot
20	47.5	27.8	345	A82664	H-acetylglutamate
21	47.5	27.8	393	A82664	conserved hypothet
22	47.5	27.8	401	A97446	hypothetical prote
23	47.5	27.8	503	A82193	hypothetical prote
24	47.5	27.8	533	A33111	sin/nuclear prot
25	47	27.5	159	G82669	segmentation prote
26	47	27.5	215	B87577	ubiquitome biosynt
27	47	27.5	284	F71015	glutathione S-tran
28	47	27.5	333	H82852	hypothetical prote
29	47	27.5	501	T46336	hydroxybenzoate oc
					hypothetical prote

30	47	27.5	938	2	C84480	hypothetical prote
31	47	27.5	4589	2	T14914	dynein beta heavy
32	46.5	27.2	240	2	D64688	probable 1-acylgly
33	46.5	27.2	779	2	T49717	related to BCS1 pr
34	46.5	27.2	957	2	A84089	hypothetical prote
35	46.5	27.2	1495	2	T31434	denisin-180 - rat
36	46.5	27.2	1607	2	T13250	hypothetical prote
37	46.5	27.2	1693	1	MMWHE	genome polyprotein
38	46.5	27.2	4957	2	T03455	AKR protein - huma
39	46.5	27.2	5262	2	T03454	AKR protein - huma
40	46	26.9	256	2	F70812	probable 1pqr prot
41	46	26.9	342	2	B64395	malic acid transpo
42	46	26.9	347	2	H64371	malic acid transpo
43	46	26.9	419	2	AH3166	hypothetical prote
44	46	26.9	688	2	A12516	hypothetical prote
45	46	26.9	698	2	T39050	hypothetical prote

ALIGNMENTS

RESULT 1

JC7607

prolactin-releasing peptide - rat

C:Species: Rattus norvegicus (Hortway rat)

C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001

C:Accession: JC7607

R:Yamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, R.

A:Title: Isolation and characterization of the rat prolactin-releasing peptide gene.

A:Reference number: JC7607; MUID:210923785; PMID:11178959

A:Contents: Spleen

A:Accession: JC7607

A:Molecule type: DNA

A:Residues: 1-83 <YAM>

A:Cross-references: DDBJ:AB040612; DDBJ:AB040613

C:Comment: This peptide induces arachidonic acid metabolite release from rat anterior

release, and stimulation of ACTH secretion from the pituitary.

C:Genetics:

A:Gene: PRP

A:Introns: 33/1

Query Match	87.1%	Score 149;	DB 2;	Length 83;
Best Local Similarity	83.9%	Pred. No. 8.5e-15;		
Matches 26;	Conservative 1;	Mismatches 4;	Indels 0;	Gaps 0;
QY	1	SRTHRSMETRTPDINPAMYASRGIRPVGRF	31	
DB	22	SRTHQSMETRTPDINPAMYGRGIRPVGRF	52	

RESULT 2

F83376

conserved hypothetical protein PA2151 [Imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83376

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mironoguch, S.D.; Warriner, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: F83376

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-664 <STO>

A:Cross-references: GB:AE004642; GB:AE004091; NID:99948163; PIDM:AAG05539.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2151

Query Match 33.0%, Score 56.5; DB 2; Length 664;

Best Local Similarity 45.8%; Pred. No. 3.4;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 4 HRHSMETRPDINPAWASRGIRP 27
DB 470 YRPNFVYTPDINP-WLQKSGRR 492

RESULT 3

hypothetical protein F15616.60 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47959

R:De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet
submitted to the Protein Sequence Database, January 2000

A:Reference number: 224480

A:Accession: T47959

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-790 <DEH>

A:Cross-references: EMBL:AL132959

A:Experimental source: cultivar Columbia; BAC clone F15616

C:Genetics:

A:Map position: 3

A:Introns: 39/1; 678/2; 698/3; 773/2

A>Note: F15616.60

Query Match 31.6%; Score 54; DB 2; Length 790;
Best Local Similarity 52.6%; Pred. No. 9.7;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 13 PDINPAWASRGIRPV 31
DB 366 PPNRPRTYGSRGQPRHGRW 384

RESULT 4

hypothetical protein RV3485c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70569

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Rogers, J.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70569

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-314 <COL>

A:Cross-references: GB:295390; GB:AL123456; NID:93261766; PIDN:CAB08708.1; PID:92104408

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3485c

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

F:46-227/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 31.3%; Score 53.5; DB 2; Length 314;

Best Local Similarity 27.5%; Pred. No. 4.2;

Matches 11; Conservative 5; Mismatches 11; Indels 13; Gaps 1;

QY 1 SRTIR-----HSMEITPDINPAWASRGIRP 27
DB 187 SRTIRMGAGVGTSAVDHKKLAADLGPSSWVAVNSIRP 226

RESULT 5

DB87559
sensory box histidine kinase/response regulator [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: DB87559

R:Nierman, W.C.; Reibold, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.;

Laub, M.T.; DeBoy, R.T.; Dodson, R.V.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A>Title: Complete genome sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: DB87559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-637 <STO>

A:Cross-references: GB:AE005673; NID:913424056; PIDN:AAK24472.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2501

Query Match 31.0%; Score 53; DB 2; Length 637;
Best Local Similarity 48.0%; Pred. No. 11;
Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 4 HRHSMETRPDINPAWASRGIRPV 28
DB 22 HRSDDLKSPALNPAIRVRLRAV 46

RESULT 6

sensory transduction histidine kinase - Methanobacterium thermoautotrophicum (strain

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: F69099

R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.

Qiu, D.; Spadafora, R.; Vlatkovic, C.J.; Mao, J.; Rice, P.; Neilling, J.; Reece, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: F69099

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-785 <MTM>

A:Cross-references: GB:AE000805; GB:AE000666; NID:92621213; PIDN:AB84680.1; PID:9262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTM174

Query Match 30.4%; Score 52; DB 2; Length 785;
Best Local Similarity 46.4%; Pred. No. 19;
Matches 13; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

QY 4 HRHSMETRPDINPAW--YASRGIRPV 29
DB 412 HRATFRRRDPDGNRYRVEYDRPRTDG 439

RESULT 7

hypothetical protein K05.ori1882 - Mycoplasma pneumoniae (strain ATCC 29342)

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999

C:Accession: S73484; S62840

R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A>Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumon

A:Reference number: S73327; MUID:97103683; PMID:8948633

A:Accession: S73484

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1882 <HIM>

A:Cross-references: EMBL:AE000017; GB:U00089; NID:91673812; PIDN:AB95806.1; PID:9167

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

Thu May 1 15:54:16 2003

US-09-446-543a-61.rapb

Page 1

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:33:06 ; Search time 19.375 Seconds
(without alignments)
138.059 Million cell updates/sec

Title: US-09-446-543a-61
Perfect score: 171
Sequence: 1 SRTHRSHMEIRTPDINPAWYASRGIRPYGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	100.0	31	10	US-09-932-161-15 Sequence 15, App1
2	171	100.0	87	12	US-10-044-592-92 Sequence 92, App1
3	158	92.4	31	10	US-09-932-161-13 Sequence 13, App1
4	158	92.4	31	12	US-10-044-592-39 Sequence 39, App1
5	158	92.4	32	12	US-10-044-592-40 Sequence 40, App1
6	158	92.4	33	12	US-10-044-592-41 Sequence 41, App1
7	158	92.4	98	12	US-10-044-592-28 Sequence 28, App1
8	158	92.4	98	12	US-10-044-592-38 Sequence 38, App1
9	158	92.4	98	12	US-10-044-592-82 Sequence 82, App1
10	158	92.4	98	12	US-10-044-592-84 Sequence 84, App1
11	158	92.4	98	12	US-10-044-592-86 Sequence 86, App1
12	158	92.4	98	12	US-10-044-592-88 Sequence 88, App1
13	149	87.1	31	10	US-09-932-161-14 Sequence 14, App1
14	149	87.1	31	12	US-10-044-592-4 Sequence 4, App1
15	149	87.1	31	12	US-10-044-592-5 Sequence 5, App1
16	149	87.1	70	12	US-10-044-592-80 Sequence 80, App1
17	149	87.1	82	12	US-10-044-592-1 Sequence 1, App1
18	149	87.1	86	12	US-10-044-592-96 Sequence 96, App1
19	149	87.1	91	12	US-10-044-592-94 Sequence 94, App1

20	147	86.0	29	12	US-10-044-592-26 Sequence 26, App1
21	128	74.9	25	12	US-10-044-592-78 Sequence 78, App1
22	113	66.1	20	10	US-09-932-161-18 Sequence 18, App1
23	109	63.7	20	10	US-09-932-161-16 Sequence 16, App1
24	109	63.7	20	12	US-10-044-592-42 Sequence 42, App1
25	109	63.7	21	12	US-10-044-592-43 Sequence 43, App1
26	109	63.7	22	12	US-10-044-592-44 Sequence 44, App1
27	105	61.4	20	10	US-09-932-161-17 Sequence 17, App1
28	105	61.4	20	12	US-10-044-592-6 Sequence 6, App1
29	103	60.2	19	12	US-10-044-592-27 Sequence 27, App1
30	88	51.5	40	12	US-10-044-592-80 Sequence 80, App1
31	57	33.3	9	12	US-10-044-592-8 Sequence 8, App1
32	54	31.6	209	9	US-10-108-915-30 Sequence 30, App1
33	54	31.6	428	10	US-09-820-155-4 Sequence 4, App1
34	50	29.2	428	10	US-09-820-155-2 Sequence 2, App1
35	48	28.1	10	12	US-10-044-592-9 Sequence 9, App1
36	46.5	27.2	240	9	US-09-895-913A-184 Sequence 184, App
37	46	26.9	417	9	US-09-738-626-4642 Sequence 4642, Ap
38	45.5	26.6	89	10	US-09-764-877-1227 Sequence 1227, Ap
39	45.5	26.6	383	9	US-10-224-446-2 Sequence 2, App1
40	45	26.3	44	9	US-09-925-299-1094 Sequence 1094, Ap
41	45	26.3	44	10	US-09-984-245-318 Sequence 318, Ap
42	45	26.3	236	9	US-09-966-262-318 Sequence 318, App
43	45	26.3	236	9	US-09-983-966-318 Sequence 318, App
44	45	26.3	236	9	US-10-143-090-318 Sequence 318, App
45	45	26.3	236	9	US-10-143-090-318 Sequence 318, App

ALIGNMENTS

RESULT 1
US-09-932-161-15
Sequence 15, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Clevelin, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods For
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 31
TYPE: PRT
ORGANISM: Homo Sapien
US-09-932-161-15

Query Match 100.0% ; Score 171; DB 10; Length 31;
Best Local Similarity 100.0% ; Pred. No. 7.7e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 SRTHRSHMEIRTPDINPAWYASRGIRPYGRF 31
1 SRTHRSHMEIRTPDINPAWYASRGIRPYGRF 31
RESULT 2
US-10-044-592-92
Sequence 92, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 92
TYPE: PRT
LENGTH: 87
ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 100.0%; Score 171; DB 12; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.4e-17;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTNRSMETRTPDINPAMWASRGIRPVGRF 31
DB 23 SRTNRSMETRTPDINPAMWASRGIRPVGRF 53

RESULT 3
US-09-932-161-13
Sequence 13, Application US/09932161
Patent No. US2002003753A1
GENERAL INFORMATION:
APPLICANT: Cliveall, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods for
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 31
TYPE: PRT
ORGANISM: Bos taurus
US-09-932-161-13

Query Match 92.4%; Score 158; DB 10; Length 31;
Best Local Similarity 90.3%; Pred. No. 5.6e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTNRSMETRTPDINPAMWASRGIRPVGRF 31
DB 1 SRAQNSMETRTPDINPAMWASRGIRPVGRF 31

RESULT 4
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Niumma, Shoji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:

SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 92.4%; Score 158; DB 12; Length 31;
Best Local Similarity 90.3%; Pred. No. 5.6e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTNRSMETRTPDINPAMWASRGIRPVGRF 31
DB 1 SRAQNSMETRTPDINPAMWASRGIRPVGRF 31

RESULT 5
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Niumma, Shoji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 40
LENGTH: 32
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-40

Query Match 92.4%; Score 158; DB 12; Length 32;
Best Local Similarity 90.3%; Pred. No. 5.8e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTNRSMETRTPDINPAMWASRGIRPVGRF 31
DB 1 SRAQNSMETRTPDINPAMWASRGIRPVGRF 31

RESULT 6
US-10-044-592-41
Sequence 41, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Niumma, Shoji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 41
LENGTH: 33
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-41

Thu May 1 15:54:16 2003

us-09-446-543a-61.rapb

Page 3

Query Match 92.4%; Score 158; DB 12; Length 33;
Best Local Similarity 90.3%; Pred. No. 5.9e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMTAGRGIRPVGRF 31
DB 23 SRAHQSHMETPTDINPAMTAGRGIRPVGRF 31

RESULT 7
US-10-044-592-28
; Sequence 28, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 28
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Murine
US-10-044-592-28

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1.9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMTAGRGIRPVGRF 31
DB 23 SRAHQSHMETPTDINPAMTAGRGIRPVGRF 53

RESULT 8
US-10-044-592-38
; Sequence 38, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 38
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-38

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1.9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMTAGRGIRPVGRF 31
DB 23 SRAHQSHMETPTDINPAMTAGRGIRPVGRF 53

RESULT 9
US-10-044-592-82
; Sequence 82, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 82
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-82

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1.9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMTAGRGIRPVGRF 31
DB 23 SRAHQSHMETPTDINPAMTAGRGIRPVGRF 53

RESULT 10
US-10-044-592-84
; Sequence 84, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 84
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-84

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1.9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMTAGRGIRPVGRF 31
DB 23 SRAHQSHMETPTDINPAMTAGRGIRPVGRF 53

RESULT 11
US-10-044-592-86
; Sequence 86, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 86
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-86

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1,9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTNRSHMETRTPDINPAMYASRGIRPVGRF 31
Db 23 SRAHNSMETRTPDINPAMYAGRGIRPVGRF 53
||:|||||

RESULT 12
US-10-044-592-88
; Sequence 88, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 88
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-88

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1,9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTNRSHMETRTPDINPAMYASRGIRPVGRF 31
Db 23 SRAHNSMETRTPDINPAMYAGRGIRPVGRF 53
||:|||||

RESULT 13
US-09-932-161-14
; Sequence 14, Application US/09922161
; Patent No. US20020037533A1
; GENERAL INFORMATION:

APPLICANT: Civeill, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods For
; FILE REFERENCE: P-UC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Rattus
US-09-932-161-14

Query Match 87.1%; Score 149; DB 10; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,1e-14;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRTNRSHMETRTPDINPAMYASRGIRPVGRF 31
Db 1 SRAHNSMETRTPDINPAMYAGRGIRPVGRF 31
||:|||||

RESULT 14
US-10-044-592-4
; Sequence 4, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Murine
US-10-044-592-4

Query Match 87.1%; Score 149; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,1e-14;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRTNRSHMETRTPDINPAMYASRGIRPVGRF 31
Db 1 SRAHNSMETRTPDINPAMYAGRGIRPVGRF 31
||:|||||

RESULT 15
US-10-044-592-5
; Sequence 5, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10

Thu May 1 15:54:16 2003

us-09-446-543a-61.rapb

Page 5

; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 5
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPRIDE
; LOCATION: (1)..(31)
; OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 87.1%; Score 149; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.1e-14;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 SRTNRHSMERTPDINPAWYASRGIRPVGRF 31
||:|||||:|||||:|||||:|||||:|||||
Db 1 SRAHQHSMERTPDINPAWYASRGIRPVGRF 31

Search completed: May 1, 2003, 14:46:06
Job time : 19.375 secs

100

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:36:28 ; Search time 84 Seconds
(without alignments)
80.947 Million cell updates/sec

Title: SEQ74-PLUS-73

Perfect score: 156
Sequence: 1 SRXHXSMEXRTPDINPAWYXXRGIRPVGRFX 33

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	93.6	98	6	OBW12
2	89	63.5	117	13	O9W624
3	69	44.2	15281	3	O09164
4	68	43.6	34350	4	O8W42
5	67	42.9	7463	16	O9Z4X6
6	66	42.3	870	4	O00308
7	66	42.3	870	4	O96C22
8	66	42.3	870	11	O9DBH0
9	66	42.3	1889	10	O94H07
10	66	42.3	4578	13	O42181
11	66	42.3	26926	4	O10466
12	66	42.3	26926	4	O8W2B3
13	65	41.7	2308	5	O9VDJ9
14	64	41.0	1171	3	O96B77
15	64	41.0	4135	6	O18977
16	64	41.0	4247	2	O9L8H4

17	64	41.0	6620	4	O96AA2	O96AA2 homo sapien
18	63	40.4	3063	2	O59497	O59497 coynebacte
19	63	40.4	3325	5	O8F510	O8F510 anophelis 9
20	63	40.4	6889	16	O8XS40	O8XS40 talstonia s
21	62	39.7	1762	2	O30480	O30480 streptomyce
22	62	39.7	2611	2	O68487	O68487 streptomyce
23	62	39.7	3546	2	O9F830	O9F830 microcosm
24	62	39.7	4133	13	O9DE12	O9DE12 gallus galli
25	62	39.7	4134	13	O8OGX4	O8OGX4 gallus galli
26	62	39.7	5825	10	O82731	O82731 vicia faba
27	62	39.7	7389	11	O91206	O91206 mus musculu
28	62	39.7	8563	2	O54297	O54297 streptomyce
29	61	39.1	565	5	O9VZY7	O9VZY7 drosophila
30	61	39.1	963	16	O9CNX2	O9CNX2 pasteurilla
31	61	39.1	967	2	O9L6A7	O9L6A7 pasteurilla
32	61	39.1	1059	5	O19370	O19370 caenorhabdi
33	61	39.1	1410	2	O9K1Z9	O9K1Z9 polyanthum
34	61	39.1	1771	17	O8TV64	O8TV64 methanopyru
35	61	39.1	3419	5	O9GP18	O9GP18 halioris tu
36	61	39.1	3638	4	O15142	O15142 homo sapien
37	61	39.1	4345	5	O9VLA0	O9VLA0 drosophila
38	61	39.1	4488	11	O9QZHL	O9QZHL mus musculu
39	61	39.1	5138	10	O9ZW94	O9ZW94 arabidopsis
40	61	39.1	6831	5	O23550	O23550 caenorhabdi
41	61	39.1	7160	5	O23551	O23551 caenorhabdi
42	61	39.1	7257	2	O9L8C7	O9L8C7 polyanthum
43	61	39.1	7257	2	O9K1Z7	O9K1Z7 polyanthum
44	61	39.1	7525	2	O9K1E0	O9K1E0 streptomyce
45	61	39.1	9477	2	O9L4X3	O9L4X3 streptomyce

ALIGNMENTS

RESULT 1

OBW12

ID OBW12

AC OBW12

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)

DE Preprolactin-releasing peptide.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Caprinae; Ovis.

OX NCBI_TaxID=9940;

RN (1)

RP SEQUENCE FROM N.A.

RA Curlew's J.D., Kusters D.H.L., Barclay J.L., Anderson S.T.,

RT "prolactin-releasing peptide (PRP) in the ewe: cDNA cloning, mRNA

RT distribution and effects on prolactin secretion in vitro and in

RT vivo."

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF450453; AAL47178.1;

SO SEQUENCE 98 AA; 10513 MW; 2A53331ED62CAAB5 CRC64;

Query Match

Best Local Similarity 93.6%; Score 146; DB 6; Length 98;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAWYXXRGIRPVGRF 31

DB 23 SRXHXSMEXRTPDINPAWYXXRGIRPVGRF 53

RESULT 2

O9W624

AC O9W624

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE

DE C-RF amide.
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Carassius.
 NC NCBI_Taxid=7957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RA Salake H., Minakata H., Fujimoto M.;
 RT "Carassius kribiaide (C-RF amide).";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB020024; BAA76662.1; -
 SQ SEQUENCE 117 AA: 12879 MW: D5DC4CB2038C2B0 CRC64;
 QY Query Match 63.58; Score 99; DB 13; Length 117;
 Best Local Similarity 39.08; Pred. No. 0.0014; 7; Indels 10; Gaps 1;
 Matches 16; Conservative 8; Mismatches 7;
 DB 35 SNAAGTVEHDLHYHVDNRSPDLDFVYGVGRVRIGRF 75
 OY 1 SRKH-----XHSNKRTPDINPARYXXRGIRPVGRF 31
 ID 009164 PRELIMINARY; PRT: 15281 AA.
 AC 009164;
 DT 01-NOV-1996 (TRENDArel. 01, Created)
 DT 01-NOV-1996 (TRENDArel. 01, Last sequence update)
 DT 01-MAR-2002 (TRENDArel. 20, Last annotation update)
 DE Cyclosporin synthetase (CYSPN) (EC 6.---).
 GS SIMA.
 OS Tolyposcladium inflatum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Clavicipitaceae; Mitosporic Clavicipitaceae;
 OC Tolyposcladium.
 NC NCBI_Taxid=29910;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 34921;
 RX MEDLINE=95094306; PubMed=8001164;
 RA Weber G., Schoergerdorfer K., Schneider-Scherzer E., Leitner E.;
 RT "The peptide synthetase catalyzing cyclosporine production in
 Tolyposcladium niveum is encoded by a giant 45.8-kilobase open reading
 frame.";
 RT Curr. Genet. 26:120-125 (1994).
 RL Cur. Genet. 26:120-125 (1994).
 CC -1- FUNCTION: THE CONSTITUTED AMINO ACIDS OF CYCLOSPORINS ARE
 ACTIVATED AS ANIONIC-ADENYLATES WITH PEPTIDE BONDS FORMED
 THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- COFACTOR: CONTAINS 11 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF CYCLOSPORINS.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN AMP-DEPENDENT
 COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC EMBL: 228383; CAAB2227.1; -
 DR HSSP; P14687; IAMU.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR000267; Asp/Glutamase.
 DR InterPro: IPR001242; Condensatn.
 DR InterPro: IPR003880; Peptide_attach.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam: PF00668; Condensation; 11.
 DR Pfam: PF00550; pp-binding; 13.
 DR PRINTS: PRO0154; AMPBINDING.
 DR PRINTS: PRO0139; ASGLINASE.
 DR PROSITE: PS00075; ACP_DOMAIN; 11.
 DR PROSITE: PS00455; AMP_BINDING; 10.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_9.
 KW Liasse; Antidiabetic biosynthesis; Phosphopantetheine; Repeat;
 KW Multifunctional enzyme.
 FT DOMAIN 15179 15219 13 x 3 AA APPROXIMATE REPEATS.

FT REPEAT 3 1086 DOMAIN 1.
 FT REPEAT 1087 2585 DOMAIN 2.
 FT REPEAT 2586 4072 DOMAIN 3.
 FT REPEAT 4073 5564 DOMAIN 4.
 FT REPEAT 5565 7061 DOMAIN 5.
 FT REPEAT 7062 8121 DOMAIN 6.
 FT REPEAT 8122 9616 DOMAIN 7.
 FT REPEAT 9617 11113 DOMAIN 8.
 FT REPEAT 11114 12185 DOMAIN 9.
 FT REPEAT 12186 13681 DOMAIN 10.
 FT REPEAT 13682 14767 DOMAIN 11 (ALA-ACTIVATING).
 FT BINDING 1060 1060 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 1060 1060 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 2558 2558 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 4045 4045 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 5537 5537 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 7034 7034 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 8094 8094 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 9589 9589 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 11086 11086 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 12158 12158 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 13654 13654 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 14729 14729 PHOSPHOPANTETHEINE (POTENTIAL).
 SQ SEQUENCE 15281 AA: 1689051 MW: E26DA7AA35324C05 CRC64;
 QY Query Match 44.28; Score 69; DB 3; Length 15281;
 Best Local Similarity 0.2%; Pred. No. 6.6e+03;
 Matches 20; Conservative 3; Mismatches 8; Indels 11650; Gaps 5;
 OY 1 SRKH-----
 ID 1351 SRKHQVTSFAVLLAARAAHFRITGSDNATIGVPSANRRPELVIGFVNTGCIIRITI 1410
 OY 5 -----
 DB 1411 DENDNESLVRYRSTTTAAQDNQVPEQVVSILMPSSRDASRNPVQLFALHGGOD 1470
 OY 5 -----
 DB 1471 LFRIOLEGTREEVYPTREYTRDIEFELVQASRLSGDITFAADPEAETTINGVSVEOE 1530
 OY 5 -----
 DB 1531 VLRGIGQDPOTPLWPLPLDGLPELERNGLIMVKTIDYRNKSVYDVVQOVRLSAEATA 1590
 OY 5 -----
 DB 1591 VIDSSRMVIAELDRSDQVAATLRORLPAAETFAVLAAPRSCAVIALFGILKAGHAYL 1650
 OY 5 -----
 DB 1651 PLDVNVPAAALRAILAEVKGKLVLLAGEPSDEGSPSVIVRIADNTSPAGHSLRDG 1710
 OY 5 -----
 DB 1711 KSRFTAGSLAVYFTSGTGKPGWVLEHGVLRVKQINILSSLPAAQTFKMAHMSNLA 1770
 OY 5 -----
 DB 1771 FDASIVEFTALLNGSLWCIDREFTILDAQALEAFIREHINI ALFPALLKQCIDTAAA 1830
 OY 5 -----
 DB 1831 TISLIDLIVYGGDRIDTADALAKALVKSEVTNANGPTENTVASTIYSTADTERVNGVF 1890
 OY 5 -----
 DB 1891 IGRAVNSGVYVMDQQLVPLGVKGLVVTGDLARGTNPALDSRFDVVIANGQLLR 1950
 OY 5 -----
 DB 1951 AYRTGRRARPRPDQGVVEFGRRDQVKVGRHRIELAEVHALSSAGVHADAVVNSQOE 2010
 OY 5 -----

Db 2011 DNOGEVAFITDADNETLQPAOSSNOVQEWESHETTAVADITADONTLGRDFTSWTS 2070
 QY 5 ----- 4
 Db 2071 MYDGTLLDKREMOBLDTHRTFLDQGAAGVLEIGTGTGVLENLQAGLAKSTIGLEPS 2130
 QY 5 ----- 4
 Db 2131 QSAVOFNKAAQTEPGEKRAQVHVGTAMDGRSLSPDLIVNSVAQYFSPREYLAEV 2190
 QY 5 ----- 4
 Db 2191 VEAIVRIPGVRIEFGDMRTYATHKDFLVAARAVHTGSKVTRSKVOQVARLEEBELL 2250
 QY 5 ----- 4
 Db 2251 VDPAFITLSESLSEIEHEVELPKNNKYNNELSYRYGAVLHRNNHNSRSIKINA 2310
 QY 5 ----- 4
 Db 2311 ESWIDPASQMDROGLAKLKENKDAESIAVFNIPYSKTYERHIAKSLADHDGDTHS 2370
 QY 5 ----- 4
 Db 2371 SIDGVAMISARKEASQCPSLVDVHDLVQLAEDAGFRVYSMAQRSONGALDVFHHHP 2430
 QY 5 ----- 4
 Db 2431 TENESRALVDFPUDYKQQAARSITNRPLOVESRRIEPAQVREQLQVLLPAYMIPARIYVL 2490
 QY 5 ----- 4
 Db 2491 QNNPLNTSGVDRKELTLAKKTAAKTPSEELVAPRDSIPATICEKFKDYLAVEVGTDN 2550
 QY 5 ----- 4
 Db 2551 FENVGSHLATKLAARLSQNAQAVNDIFDRPIADIAITQDTEHNPILTPTSTY 2610
 QY 5 ----- 4
 Db 2611 GPVQSFQAGRLWFLDOLNCGATWYIMPAVRLGFLVVSALAALEERHETLRTTE 2670
 QY 5 ----- 4
 Db 2671 IEQEGIGMOTHPAPKELAVIDVSGEESTIQILKEQTPPENLASEPFLALKTG 2730
 QY 5 ----- 4
 Db 2731 EDEHILSTVHHNAISDGWSVDIFQELQGYSAIINGHDLPAQIAPLSTQYRDPATWQ 2790
 QY 5 ----- 4
 Db 2791 IFQVAHRROLAYWTKOLADNKPABLTDKRPMLSGRAGEIPVYVGLIYEKLDPCR 2850
 QY 5 ----- 4
 Db 2851 IRQVATFVLLAFARAHRTGTBDATIGTPIANRNNRELBGLGFEVNTQCMRTVDY 2910
 QY 5 ----- 4
 Db 2911 EDSFETLVHQBRETLAANADVPFEQIVSNILPQSSDTSRNPVLQVLFALHSQONLCK 2970
 QY 5 ----- 4
 Db 2971 VRLGIEELIETAEETRPDLIEHLYQEAERLNGSIVAADELEVETITQSVITITQGLQ 3030
 QY 5 ----- 4
 Db 3031 KGLGEPDMFVASKALDGLSLRSTGLHPQOTDPCDASVQIFKQVAVNPDIYAVRD 3090
 QY 5 ----- 4

Db 3091 ESTRLSYADIDRKSQDVACMLSHRGIAPEFVALLAPRSCETIVALLSVLRANLAVPLD 3150
 QY 5 ----- 4
 Db 3151 VNPASRLNILEVSGSMVLVGAETPIEGMAEETIRITETLADATDDINGLASQ 3210
 QY 5 ----- 4
 Db 3211 PTASLAVVITSGTGRKGVNVEHRTVRLTKQNTITSKLPSEFHAIHSLAFDASV 3270
 QY 5 ----- 4
 Db 3271 WEYFTLLNGCLWCIDYFTLLESTALEKVPFQRYVWALLPALLKQCLDNPALVTL 3330
 QY 5 ----- 4
 Db 3331 SVLTIGDRLASDAKANGLVOTQAFNAVYPTENTVSTYPIADEPTEINGVPIGAVS 3390
 QY 5 ----- 4
 Db 3391 NSGAFVMDQNOITPPGANGELIYTGDLARGYTSLSNTERFINVDIDEGVARYRTGD 3450
 QY 5 ----- 4
 Db 3451 RVRTRPKDQIEFFGRIDHQRIRGRIRIEPAVEYALLSHDLVTDAAVYTHSQENDLEM 3510
 QY 10 ----- 9
 Db 3511 VGFVARAVADVEDSSNOVQEWQTHFDSIAYADITTDQSLGRDFTSWTSMQDGLIK 3570
 QY 10 ----- 9
 Db 3571 KSOQEWLDDTRSLDQSPGHVLEVGTCMVLFNLGRBSGLQSTYGLRPSBATAFY 3630
 QY 10 ----- 9
 Db 3631 NKAASFQLEDRIEVEGTATDIDLGLDMLAGLVVNSVAQYFSPQDYLAQLVRDLTK 3690
 QY 10 ----- 9
 Db 3691 VPGEIRIFFGDMRSHAINRDELVAARAVHALGDKATKAEIQREVVMESEDELLVDPATF 3750
 QY 10 ----- 9
 Db 3751 TSLTQVENIKHEVELPRKMATNELSSRYAAVLHVNDLAKPAHKVSPGAVDPAATKM 3810
 QY 10 ----- 9
 Db 3811 DRDALIRLLRGTKISDIALANIPNSKTIIVERTICESVYDLGDAKDSNDREVSWLSARS 3870
 QY 10 ----- 9
 Db 3871 NAVKVASLADLVDAQAGRVEISCARQMSONGALDAVFHNLGSPQSSHVLIDELT 3930
 QY 10 ----- 9
 Db 3931 DHGRPEALTNHPLHRAQSRVERQIRERLQTLIPAYMIPQIMVLDKPLANNKQVDR 3990
 QY 10 ----- 9
 Db 3991 KQVQRAQTVKAKQVSAFVAPRTEIERVLCQEFSDVLCVDIGITMENFEGLGSHMATK 4050
 QY 10 ----- 9
 Db 4051 LAARISRLTEHVSKEIFDHRVCDLVLIQGSAPHDPIVSTKYTPQSFQAGRLM 4110
 QY 10 ----- 9
 Db 4111 FLQDLNFGATWYLMPLAVRLNGANNVHALTAALLAERREHLLRTTEYEQNGVGMQVNP 4170
 QY 10 ----- 9
 Db 4171 VVETIRIIDISSNGDGYLPTLKRQOTAPFHLETEPGRRVALLRGPQDYILSVVAKHII 4230

```

QY 10 ----- 9
Db 4231 SDGMSVDVTLFQELGQFYSTANGEDPLSQTPLPIHYRPAIMQKPTQESHEHQLOTV 4290
QY 10 ----- 9
Db 4291 VEQVDSAPAEILLDLPSPSILSGQAGMSVTLEGALYKNEECVHRVTSFVLLAAL 4350
QY 10 ----- 9
Db 4351 RAAHYRLTGSSEATIGTPIANRRPELEQITGFVNTQCRITVNEDETFESLWQVNST 4410
QY 10 ----- 9
Db 4411 ATAFAHQDVPEKIVSTLLPGSRDASRNPVQMFVHSQKNLGELAKENAHSEVVPTE 4470
QY 10 ----- 9
Db 4471 ITTRFDLEFHLFOODKRLBGSILYSTDLFEAVSVQSLISVQELRGLNGPDPVPISTLP 4530
QY 10 ----- 9
Db 4531 LQGGIVDQROGLDVQKTEYPRDSSVVDVFEHQVSINPDSIALHGSEKLSAQDRBS 4590
QY 10 ----- 9
Db 4591 DRYAFMLRHSFSSDTLIANVAPRSCETIIATGILKANLAVLPDYKAPANIDAIVSS 4650
QY 10 ----- 9
Db 4651 LPGNKLLILGANVTPPKIDBAIDVPPIRDTFTLLDGLQDPTTERPSAQSLAVAMT 4710
QY 10 ----- 9
Db 4711 SGTGRKGVWQHRNIVRLYKNSNVYAKQPAARIAHISNLAFDASSWETIAPLANGA 4770
QY 10 ----- 9
Db 4771 IVCADYFTTIDPQALQETQEBHEIRGAMLPSSLKQCLVQAPDMISRLDILFAAGDRSS 4830
QY 10 ----- 9
Db 4831 VDALQORLWSSGVFNAYGPTENTILSTIYVAENDSFVNGVPIGSAVNSGAVIMDKQ 4890
QY 10 ----- 9
Db 4891 QLVYAGVMSLVYTGDLANGYMDPKLDADRFTQLTVNSGEQVRAVYTGDRVAYRPRDPQ 4950
QY 10 ----- 13
Db 4951 IEFGRMDQOIKIRGRIEPAEYDQAFNDGFVEDVAIVIRTEMDPEKAVAYTAKGN 5010
QY 14 ----- 13
Db 5011 SARBEATQTQEGEAFBEGGAVANTIEPISSEALGYDFKQNTSKYDSTELDKDREHNLN 5070
QY 14 ----- 13
Db 5071 DTKMSILDGRKAGVLEVGVTGKIMFNILGRSGILERYIGLEPAPSAEYVNNAAKSPFG 5130
QY 14 ----- 13
Db 5131 LAGRAEVHVTADVGTQGLTSDMAVINSVAOYFPTPEYLAETIKSLVOYPGKRITYG 5190
QY 14 ----- 13
Db 5191 DMKSHANNRDEFAARAAYSLADNASKDVRVQKMELEKEBELLYDPAFFALASQLODR 5250
QY 14 ----- 13
Db 5251 IQHVEILPKRKATNELSSYAAVLHISDEPLPIYKIDPEAMINFGSRLTREALAQVL 5310

```

```

QY 14 ----- 13
Db 5311 KENNAESVAISNIPYSKTVVERHIVSLDQEDANAPESMDGSDWISAVTRAQCHTL 5370
QY 14 ----- 13
Db 5371 SASDLEFIADAGFVEVSNARQSHOGALDAVHRLKPALEDSESVLLKEPPTHQGNPLK 5430
QY 14 ----- 13
Db 5431 SLTNOPPLPAQSRRAELLIREGLOTLPPYKIPSOITLLIDRMPINANGKYDRRELBARRAK 5490
QY 14 ----- 13
Db 5491 ITOKSPVEDIVPRNSYEATVCKGFTDVLGEVGTIDNFEHLGHSIMATKLAARLRQ 5550
QY 14 ----- 13
Db 5551 LNTRISSHVDVDPVADLAAVIQRNSAPHEPIKPADYTGVPQSFAGRLMFLDQLVG 5610
QY 14 ----- 13
Db 5611 ATWYAMPGLIRLHSLRYDALATATISALQHRHPLRTFHEDEGVGVVQDHRPKDLRI 5670
QY 14 ----- 13
Db 5671 IDLSTQPDVAVLAKHQTLTFLATEPGNRVALIRLGEERHILSYMHIIISDGSVE 5730
QY 14 ----- 13
Db 5731 VLFDMHNFYSSALQOQPMEOILPPIQYRDPAMQOTERQVABHQRLDYTEHLADS 5790
QY 14 ----- 13
Db 5791 TPABLLTDLPRPSILSGRANLPLTIEGRILDKLNAFCRVHQATPEVILLALRAAHYRL 5850
QY 14 ----- 13
Db 5851 TGAEDATIGTPIANRRPELEKMGFEVNTQCMRIAEENDFESLVRVRSATSAFAN 5910
QY 14 ----- 13
Db 5911 QDVPESTIVSSILPGSRDASRNPVQVILAVASQODLKLLEGLRDAVDSATSTRDV 5970
QY 14 ----- 13
Db 5971 EFELFEHADRLSGVLYAKELKLTIESVVSFELETLRRLDQPLTBLAVPLDVGGE 6030
QY 14 ----- 13
Db 6031 IASKGLDVPRDYPRDANIVEVFOQHYRANPPDAIYVANTSILTYAQLDQSDRLAINTL 6090
QY 14 ----- 13
Db 6091 SRHMAPETLVGLAPRSCETIIAMFGIKKANLAVLPDINSAPARLSILSAVGNELV 6150
QY 14 ----- 13
Db 6151 LLSGVTAPQENPEYEAAGIOETIAGTGLDGTQGSNARPSATSLAYIYFISGSGPKG 6210
QY 14 ----- 13
Db 6211 VAVERSVTRLAKPSNVISKLPOGARVAHLANIASDASIWETATLLAGATLYCIDIHTV 6270
QY 14 ----- 13
Db 6271 LDCRTLKEVERESITVTVLMPALLKQCYAIEPETIAHLDLYTGDRVGHDAARRASL 6330
QY 14 ----- 13
Db 6331 VKIGESGYGPTENTVISTIYEVDADEKFEVNGVPIGKTVNSGAVYKDRNQLVPSGVVG 6390
QY 14 ----- 13

```



```

Db 6391 ELVYTGDLARGYTDPSLANKRPIYITVNGESIRATNTGRVRRKRPNDLQIEFFGRDQO 6450
QY 14 ----- 13
Db 6451 VKIAGRIBEPGEVESALLSHNSVODAAVVICAPADDSGAENVAFVARNTEDEDEQEEE 6510
QY 14 ----- 13
Db 6511 AVDOVQNETHEFETAAYS EWKDI NQSEVGNDFMGWTSMDGSELDKTDMEHENDIWRM1 6570
QY 14 ----- 13
Db 6571 LDAREPHVLEIGTGGMVFNKACPCGLOGYGVFEPSKSAQFVNDAAQSPALADGNS 6630
QY 14 ----- 13
Db 6631 IYHVGTATDINKAGPIQPRLVJINSVAQVPEPTEYLFREVEALVQIPSVERTIVEGDKRTN 6690
QY 14 ----- 13
Db 6691 AINRDFVARSALHTGKANKRLVRQMIYELANEELLTDPAFTSLTRIGEXIKHYE 6750
QY 14 ----- 13
Db 6751 ILPKMKATNELSKYRRAVIAVHGRSREOSTINQVSPNMIIDFAADGDLNQT1INLLKEH 6810
QY 14 ----- 13
Db 6811 KDAQVAIGNLPYSKTIYERFVNKSLSEDMEGNSLDGSAMVAARMAQSCPSLDAM 6870
QY 14 ----- 13
Db 6871 DYKRIAGEAGYQVEVSKARQWSONGALD1FHFEPPKGCARLTLEPTDYEGRNVTILT 6930
QY 14 ----- 13
Db 6931 NRPLNSIOSRLGTQIREKLTLLPYMLPSRLNVLDMQVYNNNGKIDRELVRAIYAP 6990
QY 14 ----- 13
Db 6991 KPSAATRAVAPNEIEALINDEFEDVLTGEVSLDNFPLDGHSLMAYTKLAARSHRDA 7050
QY 14 ----- 13
Db 7051 HISIKOVFDQPVLAADLAASIQRESAHEPTPORPYTGPAQBSFAGRLMFLDQNLGATW 7110
QY 14 ----- 13
Db 7111 YLMPALIRINGOLRVAALSMALFALERRRHETLTTFEESDGVGVQIWEARNSDLAVHDV 7170
QY 14 ----- 13
Db 7171 STGDDEYILEVLRREQTVFPLDSEPGWRYCLVKTGEEDHVLSTVNHIIITDGSVYDILR 7230
QY 14 ----- 13
Db 7231 GELGQYSAALRGDPLMANPLPIQYRDPAMQRAKQVEHQKOLGYSKQYDSTPA 7290
QY 14 ----- 13
Db 7291 ELLTDLPRPSILSGRAGSVDTIEGSVYALQSFCRTSRVTFVLLVFERIAHFRLTAV 7350
QY 14 ----- 13
Db 7351 DDAITGTPIANRRRPELETIVGCFVNTQCMRIS IADODNFBGLVROVRNVAATAAYANDV 7410
QY 14 ----- 13
Db 7411 PERIVSALVPGSRNTRSRNPVLQVLMFAVQSVEDYDQVRLEGLIESVMMPEASTRDMETH 7470
QY 14 ----- 13

```

```

Db 7471 LVPQOKLTGSVLYSSDLEQGTIONFVD1FQECIRSVLDQPLFPISVLPFSNAISNES 7530
QY 14 ----- 13
Db 7531 LDLEMPSTDIYPRODRIVDLFREQAICPDSTAVDSSQLTYAQLDQSDRVAAMLER 7590
QY 14 ----- 13
Db 7591 HMPAESLVSPRSCETIIAYFGIMKANLAVLPDYTAAPDARLAILDYEGERLLIG 7650
QY 14 ----- 13
Db 7651 AGVPOPGIQLPRLSTAYIAEALSHATVDYTSIPQPSATSLAYITFTSGTGKPKGVNLE 7710
QY 14 ----- 13
Db 7711 HRGIVRLVDTNWNVPFSGSALPVSHSNLAMDATIEITYAVLNGGTVCIDRDTMID 7770
QY 14 ----- 13
Db 7771 IALNSTFRKENVRAAFETPAFLQOLAFETPELVANLEILHTAGDRDPGDANLAGTKAX 7830
QY 14 ----- 13
Db 7831 GGLFNVLGHTEHTAVSTIYPVVGEETFYNGVPGRGISNSHAYIIDRQKLVPAVGMEL 7890
QY 14 ----- 13
Db 7891 ILLGDSVARGYTDNALNDRFYID1NGKSTSYTGCKAKRRPDQGLEFFGRDQNVK 7950
QY 14 ----- 13
Db 7951 INGVRIEAGEVELITLDHKSVLATVAVVRPPNGDPEMIAFTIDAEDDQVTHKAIYKL 8010
QY 14 ----- 13
Db 8011 OGILPAIYIPSHLVITLDQPTVDNGKVDKDLALPAQTQVKRRSTAAVRPPREVEAVLC 8070
QY 14 ----- 13
Db 8071 EEXSNLEVEVQITDGFPLDGHSLMAYTKLAARLSROLNTRSVKXVDFDQILADLADI 8130
QY 14 ----- 13
Db 8131 RRGSRHDPITATPYTGVEQSFAGRLMFLQNLGASWTAMPFATRMGPLOTKALAV 8190
QY 14 ----- 13
Db 8191 ALNALVHRHEALRTTFEDHOGVGVQIOPKSSODLRIILSDAVDDTAIYLAALKREQITA 8250
QY 14 ----- 13
Db 8251 FDLTSEPGWRYSLRGDDYILSTVNHIIISDGVTVDLRQELGQYSAALRQGEPLSQ 8310
QY 14 ----- 13
Db 8311 AKSLPIQYRDAVWQROENQKEQAKOLKYWSQOLADSTPCFPLDLPBPSILSGEADAV 8370
QY 14 ----- 13
Db 8371 PMVLDGTVYQLLTDECRTHOYTSFSVLLAARTAHYRLTGTLDTAVGPIANRRRPELEG 8430
QY 14 ----- 13
Db 8431 LIGFEVNTQCKRMA1SETETFEESLVQVRLTTEAFANODVPFQIVSTILPGSRDTSRN 8490
QY 14 ----- 13
Db 8491 PLVOYMFALQSOODIGRIQLEGMTDEALETPLSTRDLEVHLROEVKLSGSLYSTDLF 8550
QY 14 ----- 13
Db 8551 EVETIRGIYDVFLEILRGLQPKORLAMPITDGIKRLNDQGLITVAKPAYPRESSVID 8610

```

```

QY 14 ----- 13
Db 8611 LFKOYAAAPDALAVMDSSTLTYYADLDQSNLHAWLQORNAAPETLVAVARPSCL71 8670
QY 14 ----- 13
Db 8671 VAFGLVKANLAVLPIDVNAAPARIBAILSAVPGHKLVIYOHAGPELGLTMADTELVOJD 8730
QY 14 ----- 13
Db 8731 EALASSSSGDHEQ1HASGPPTAVSLAVYMTSGSTGRPKGMDHRSITIKLVKNSDVATL 8790
QY 14 ----- 13
Db 8791 PTPVRANVSNLAFDISVGEITYTALLNGSTLVCLDLYTLTLDISKILXNVFEAVNNAKFT 8850
QY 14 ----- 13
Db 8851 PVILKQCLGNMRAIISRLSYLFNVGDRLDADHAAVAAAGLIDAVYNAVYPTENGQSTWY 8910
QY 14 ----- 13
Db 8911 KVDVNEPFGVCP1GKSTNSGATVMDGNQOLVSPGVMEIYVTGCLARCTDSALDED 8970
QY 14 ----- 13
Db 8971 RFVHTIDGSENIKAVRTGDRVRYRPDEPIEFGRMDQVKIRGHIEPABVEHALLGH 9030
QY 14 ----- 13
Db 9031 DLVHDAVAVLRKANQBPENIAFTSQEDETIQHESNKOVOGMEHFDVSRYADIKDD 9090
QY 14 ----- 13
Db 9091 TSTFGHPLGNTGKVDIPLVNEKMEIDETIASLIDNRPPGHILEIGAGTMIISNLG 9150
QY 14 ----- 13
Db 9151 KVDGLKRYGLDPAPSAITFVNEAVKSLPGLAKARVLTALDGLDNEIOPELVYI 9210
QY 14 ----- 13
Db 9211 NSVAQPTSEYLIKVKAVVEVSVKRVFGDIRSQALNRDPLAARAVALCDNAKSKQ 9270
QY 14 ----- 13
Db 9271 IREKIALESEBELVDPAFVSLBSQLPNIKHEVILPKIMKATNELSYRAAVIHIS 9330
QY 14 ----- 17
Db 9331 HNEEQULLIJDIDPTAWDEPAATQKDSQGLRNLLQGRDDVMIAVGNIPYSKTIVERHIM 9390
QY 18 ----- 17
Db 9391 NSLDQDVNSLDTGKSLDARSAAICTSPDAPALIQIAKEGPRVELSNARQSONGAL 9450
QY 18 ----- 17
Db 9451 DAVHRLATDANCERSVLYHFTPDHGRQLRLLTNRPLOQASRTIESOVFAIDOTALP 9510
QY 18 ----- 17
Db 9511 AYMIPIRIYLPQHPFNANGKVDKOLARRAOVAKKRAVSARVAPRNDTEIYCEBYAD 9570
QY 18 ----- 17
Db 9571 ILGTEVGLTDFEDMGHSLMATKLARLSRLDTRVTVKEVFDKPYLADLASIEOGST 9630
QY 18 ----- 17
Db 9631 PHLPJASSVYSYSGVEQSYAQORLWFLDQFNLNMTWYHMSLAMRLGLPDLNDALDVALRAL 9690
QY 18 ----- 17
Db 9691 EGRHETLTTEFAQKXKDIGVQVVEHAGMRILKVLJLSDKNEKEHMAVLENQMRPETLASE 9750
QY 18 ----- 17
Db 9751 PGMKGLARIGPTETIISLVNHNHSDGWSVDILROELGOYSAALGRDPLSQVPLPI 9810
QY 18 ----- 17
Db 9811 QYRDFAAQKMAQVAHEHROLATWENQALDSTPGBELLTDFPRPQPLSGKAGYIPYITEG 9870
QY 18 ----- 17
Db 9871 PVYKILFKSERQVTLFESVLLTAFRAHFRLTGAEDATIGTPIANNRPELEHIIGFV 9930
QY 18 ----- 17
Db 9931 NTQCRLLDGTSTFESLVQHVRSVANDAYSNOIPEFRIVSALLPQSRDASRSP1QIM 9990
QY 18 ----- 17
Db 9991 FALHSQPLGNITLLEGLHEHRLPTSVANRPFDMFHLPOEPKLSGSILFADELPOPETIN 10050
QY 18 ----- 27
Db 10051 SVYVFOELIRGLDQPOVSISTMPLDGLTDLKLGLEIESNFPDYSVDVFRQV 10110
QY 18 ----- 27
Db 10111 AANPAVAVDSYTSMTSLDQKSEQIAANLHAQGLPESELICVMAPRSPETIVSLFGI 10170
QY 28 ----- 27
Db 10171 LKAGVATLPLDVNSPARIQPLTSEVEGRVYLLGSGIDMPQSDRMVETARIQDILINT 10230
QY 28 ----- 27
Db 10231 KYERSDPSRPSATSLAVIETSGSTGRPKGVMIEHRLILVKQSNVSQLPQDLNMAH 10290
QY 28 ----- 27
Db 10291 ISNLAFLASIWELFTAILNGCALICIDYFTLLDSQALTTTEKARVNAITLAPALLKCL 10350
QY 28 ----- 27
Db 10351 NHAPTLFEDLKYLTIGDRLDATDAKIQALVKGTVNAVGPTEYVASTYTLTDESHY 10410
QY 28 ----- 27
Db 10411 ANGVPIGNVSSGAVINDQORLVPQGVGELVVSQGLARGTYNSTLNDREFVDIYN 10470
QY 28 ----- 27
Db 10471 DOKARVKTGDRTRRPKDGSTIEFGRMDQVKIRGHVPEAEVBOAMLGNKAIHDAVY 10530
QY 28 ----- 27
Db 10531 VOAVDQETEMIGFVMSADREFSEGEERTINQVEMEDHFEISTAVAGIEAIDQATLGRDF 10590
QY 28 ----- 27
Db 10591 TSWTSYNGNLIDKAKMEEMLDOTNQSILDKEDARCAEIGTIGVMVLEPNLPRKNDGLESY 10650
QY 28 ----- 27
Db 10651 VGLPBRSAALFVDKAADQFPLOGRQTQIIVGTARDIKLVKDFHRDVAVINSVAQYPSR 10710
QY 28 ----- 27
Db 10711 SYLVQIASLEIHMTSVKTIFFPGDMRSWATNRDPLVSALYTLGDKATQDQIROEVARLEE 10770
QY 28 ----- 27

```

```

Db 10771 NEDELLVDPAPFTSLTSGPQKVKHVELLPKRMATSNELSSRYAAVAILICRDEGRNRY 10830
QY 28 ----- 27
Db 10831 GRVHSEVENAMIDFASSGMDRLALVOMLDERBDKTPVALIGNIPHSNTINERHFTTSLDT 10890
QY 28 ----- 27
Db 10891 EGEGLADSLDSAMQSAKAMARCPOLSVTELVEIQAGFREVESWARQSRQHGALD 10950
QY 28 ----- 27
Db 10951 VVPHIEDDRGRVLINEPDPFERLPSTGITSRPLDRIONKREESQIREQLQTLPPYM 11010
QY 28 ----- 27
Db 11011 VPSRIVLERMPLNANSKVDEKELARRAFLQTIKPSATRYAPRNDIEAVLCDEFQAVIG 11070
QY 28 ----- 27
Db 11071 VTGVMONFEBELGSHLMATKLAARSLRDLTVRSKDIENOPILQDLADVOTGSAPHE 11130
QY 28 ----- 27
Db 11131 AIPSTPYSGPYEGSFGRLMFLDQNLNMSVTHMPLASRLRGPRLREALOALATTIAR 11190
QY 28 ----- 27
Db 11191 HESLRTFEEBODGVPOIVNAAKMKOLITDVGTEADAYLAALKEQDAFDMIAEPCMR 11250
QY 28 ----- 27
Db 11251 VALLRLGPDDHVLSTVAMHHTISDGWSVDILRQELQGLYSNASSQAPALPTQYBDPAIMOK 11310
QY 28 ----- 27
Db 11311 ODSQIAEHQOLNWKRLQVNSKPABELLADFTPRKALSGDADYIPIEIDQYQNLRSFC 11370
QY 28 ----- 27
Db 11371 RARHVTSTVALLAAFRAAHRTLGAEDATIGSPIANKRPELBGLIGCFVNTQCLRIPIVK 11430
QY 28 ----- 27
Db 11431 SEDTFTLVQARETAEADQNDVPERIVSSKVASSTOTSBNPLVQWFAVHSQHDIG 11490
QY 28 ----- 27
Db 11491 NIREGVEGKPVSMASSTREDAEMHLEFEDQMLGNVVFESKDLFESBTIRSVAVAFQETL 11550
QY 28 ----- 27
Db 11551 RRGIANPHANLALPLTDGLPSRLSLCLQVNPDPYRDAASYIDVFEQVASIPKSIAYID 11610
QY 28 ----- 27
Db 11611 ASSQLYTELDERSQGLATRLARQYVPEELVGVLAFRSCETIIAFLGIIKANLALPLD 11670
QY 28 ----- 27
Db 11671 VNPAGRIETILSLPGNRLILIGSDTOAVKLHANSVRFTRISIDALVESGSPTEELSTR 11730
QY 28 ----- 27
Db 11731 PTAQSLAYVFTSGSTGVKGVNHEHRTIRLVKNSNVAKOPAAAIAHLSTNIAFDASS 11790
QY 28 ----- 27
Db 11791 WEIYAPLNGVIVCIDYTTTIDIKALEAVFKOHTIRGALMLPALIKQCLVSAPIWISL 11850
QY 28 ----- 27

```

```

Db 11851 ELIFAGDRLSODALLARNAVSSGYVNAVGPTEENTVLSITHNIGENEAFSNGVPIGNV 11910
QY 28 ----- 27
Db 11911 SNGCAVMDQNOOLVSAVIGELVVTGDGLANGYTDSKLRVDFIYITLDGNRVAVRFG 11970
QY 28 ----- 27
Db 11971 DRYRHRPKDQIEFFGRMDQIKIRGRHIEPAVEQALARDALISDAVITQLTDBEEPE 12030
QY 28 ----- 27
Db 12031 LVAFSLKGNANGTKNGVNGVSDQEKIDGDEQALLMKNIRHNLQRLFTYMIPSRIHV 12090
QY 28 ----- 27
Db 12091 DQLEPVANCKIDRNEIARQAATPRTSSVTVAPRNDIETICKEFADILSRVGTIDN 12150
QY 28 ----- 27
Db 12151 FEDIGSHLATAKLAARSLRDLTVRSVDPDTPVYQGLASIQGSTPHHAIPLSHS 12210
QY 28 ----- 27
Db 12211 GPVQGSFAGRLMFLDRFNINAAVYIMPEVRLRGPIRVDALOTALRALEERHELLFTTE 12270
QY 28 ----- 27
Db 12271 EBDGVMQIVHSPRMDICVVDISGANEDAKLEEQGAPNLSTEVAMRVALFKAGEN 12330
QY 28 ----- 27
Db 12331 HHILSIYMHHTISDGWSVDIFQOELAQFSVAVRGHDPISQVKPLPIHYRDPVAVQRODK 12390
QY 28 ----- 27
Db 12391 QVAVHESQLOVTEIOLADSTPAEILSDFNREPLSGEAGTVIYEDVEYERLSIFCHNH 12450
QY 28 ----- 27
Db 12451 QVTSFVLLAAFRVAHYRLTGAEATIGPIANKRPELEDLIGFTVNTQCKRLAEHD 12510
QY 28 ----- 27
Db 12511 NFLSVRRVSTMASAFENQDVPFERIVSALLPGSRDSRNPVLQMFVHSQRRLGKQ 12570
QY 28 ----- 27
Db 12571 LBGIGEPPTPTATTTRDVEERHLEQDKGLAGNVVPAADLEAATIRSVSEVFEHILRG 12630
QY 28 ----- 27
Db 12631 LDQPIAISTWELVNDGLAALNSRNLPAVEDIEPDREASVVDVQYQVYANPDALAVTD 12690
QY 28 ----- 27
Db 12691 TSTKLTVELDQSDHVAAMLKSKQRLPAESIVVYLAPRSSEETIVACIGILKANLALPLMD 12750
QY 28 ----- 27
Db 12751 SNVPEARQALISEIPGEKPVILGAGVPIPNKTDADVAMVFIIDIVASKTDKSYSPGTRP 12810
QY 28 ----- 27
Db 12811 SASSLAYIVFTSGTGRKGVNHEHRTISLVKONASRIPQSLMAHANSNLAFASWEI 12870
QY 28 ----- 27
Db 12871 FTTLANGTLFCISTYFTVLDSKALSAFSDHRINITLIPALLKQCLADAPSVLSLESL 12930
QY 28 ----- 27
Db 12931 YIGDRLDGADATKVKDLVYRKATNAVGPTEASVNSTIYITEHETPANGVPIGTSLQPKS 12990

```

QY 28 -----YGRF 31
 Db 12891 KATIMDQDQPLVAGVAGELVAGDGLANGYIDPISLNTGRF 13031
 RESULT 4
 ID 08WZ42 PRELIMINARY; PRT: 34350 AA.
 AC 08WZ42;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE titin.
 GN TTN.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20309627; PubMed-10850961;
 RA Feilberg A., Trombitas K., Heil W., Cazorla D., Fougereuse F.,
 RA Cantner T., Kolmerer B., Wilt C., Beckmann J.S., Gregorio C.C.,
 RA Granzier H., Labelt S.;
 RT "Series of exon-skipping events in the elastic spring region of titin
 RT as the structural basis for myofibrillar elastic diversity.";
 RL Circ. Res. 86:1114-1121(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21573839; PubMed-11717165;
 RA Bang M.L., Cantner T., Fornoff F., Geach A.J., Gotthardt M.,
 RA McNabb M., Wilt C.C., Labelt D., Gregorio C.C., Granzier H.,
 RA Labelt S.;
 RT "The complete gene sequence of titin, expression of an unusual -700
 RT kDa titin isoform and its interaction with obscurin identify a novel
 RT 2-line to I-band linking system.";
 RL Circ. Res. 89:1065-1072(2001).
 DR EMBL: AJ277892; CAD12456.1;
 DR InterPro: IPR000283; Cytochrome receptor_2.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000577; EGGL_kin.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001092; HUH_basic.
 DR InterPro: IPR003598; IG.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003066; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR InterPro: IPR002016; Peroxidase.
 DR InterPro: IPR004168; PPAK_motif.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00044; fcs; 132.
 DR Pfam: PF00047; Ig; 146.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF02818; PPAK; 53.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00060; FN3; 133.
 DR SMART: SM00409; IG; 167.
 DR SMART: SM00408; IG_C2; 148.
 DR SMART: SM00406; IG_V; 23.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TYRc; 1.
 DR PROSITE: PS00933; EGGY_KINASES_1; UNKNOW_1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOW_1.
 DR PROSITE: PS00290; IG_MHC; UNKNOW_1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOW_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; UNKNOW_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; UNKNOW_1.
 SO SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;

Query Match 43.6%; Score 68; DB 4; Length 34350;
 Best Local Similarity 0.1%; Pred. No. 3.1e+04;

Matches 19; Conservative 4; Mismatches 8; Indels 31769; Gaps 5;
 QY 1 SRX-----4
 Db 2505 NRTMASDGPYKLVGRVETNCNLSVEKIIIRGLRDLCTETQNVPEVELSHSGIDL 2564
 QY 5 -----4
 Db 2565 WFKDKKIPSSKVKIEAHGIYKLVNLNMMKDEGKYTFAGENITSGKLVAGAIK 2624
 QY 5 -----4
 Db 2635 PLTDQVAESQEAPECEVAPNDKGLRDKHLPLTNIRSDGKRRLLIAAKLD 2684
 QY 5 -----4
 Db 2685 DIGEYTVKANSKISAKLVENAKIKTLKMLVTETQDAFTVELHPNKGWTKNG 2744
 QY 5 -----4
 Db 2745 VVLESNEKVAISVKGITSLRIKNCALVDESYYGRLGLGASARLNEYVKIIRKPDV 2804
 QY 5 -----4
 Db 2805 TALENATVAEVSVDTPVPKWEHKNVELRPSKXRLVSEKRYHKLMLONISPSDAGEY 2864
 QY 5 -----4
 Db 2865 TAVVGLQCKAKLFEYTLITFTMKNIETPKTASPECEVSHFNVPISMLKNGVEIEM 2924
 QY 5 -----4
 Db 2925 EKKFYVQCKLHLLIMTSTEDSAEYTFVGNQVSATLVTPIMTSMKDINAERKD 2984
 QY 5 -----11
 Db 2985 TTFEYTVYBGSITKMLKNGVEISTDCQKRTKLLHSLNIRVHNGDAADTYFVAGK 3044
 QY 12 -----11
 Db 3045 AVSTATLYEAKHIERKHINDIKVLEKRAPECEVSEPDITVOMKQDELQITDRIK 3104
 QY 12 -----11
 Db 3105 IOKERYVRLILPSTFMSDAGTYVAGNVSTANLFEGRDVRIRISIKKEVOYIKORA 3164
 QY 12 -----11
 Db 3165 VVEFEVNEVDVAHNTKDSIELNPOVERHKYVERRIHRMISFTRQSDAGEYTFVAGR 3224
 QY 12 -----11
 Db 3225 NRSVTLVNADEPPQVLOELQPVTVQSGKPARFCAVISGRFPKISWYKEBOLLSTGFK 3284
 QY 12 -----11
 Db 3285 CKFLHDGQGYTLILTRAPEDAAYTCEAKNDYGVATVSASLSEVPEVSPDQEMPPYP 3344
 QY 12 -----11
 Db 3345 PALITPLDQVTSQGPAPFQCRVSGTDLKVSYSNDKIKRPFRRYQFEDTVQLEIA 3404
 QY 12 -----11
 Db 3405 EAVPEDEGYTVFVANSVAGVSTANLSLAEESILHERIEQIEEMKEKFSSEPLSAEE 3464
 QY 12 -----11
 Db 3465 EGIHSAELQSKINTLELLSPVSYTFPDEKKBGTGTFIKVSNADISMGDAVATSV 3524
 QY 12 -----11
 Db 3525 TVIGIPKPIQWFGVGLTPSADYKVFVDDGHSLLIFTLDEDEGEYTCMASNDYGR 3584

```

QY 12 ----- 11
Db 3585 ICSATLINSKGBCHKDTESESAAVAKSLKLGCPHFLKELPIRCAGLPAIFETV 3644
QY 12 ----- 11
Db 3645 VGERAPYTWKENCOLCTSVYTTIINPNNGSCTIVNDPORDSGLYICKAENMLGEST 3704
QY 12 ----- 11
Db 3705 CAELVLLEPDMUDVPCRAKSTPEAPEDEPQTPKGPAYEALDSQELATFVKDTLK 3764
QY 12 ----- 11
Db 3765 AALTEENQOLSTEHIAKANELSOLPLGAOELQSTLEQDKLPESTREFLCINGSIHPQ 3824
QY 12 ----- 11
Db 3825 PLKEPSPMLQLOIYOSOKTFSKEGLMPEEPETOAVLSDTEKIFPSAMSTEQINSLTVER 3884
QY 12 ----- 11
Db 3885 LKTLIAPEGNTPQSIIEPWHSYLTSVAEEVLSPKETVSDTNRQRYTLQOEAQAL 3944
QY 12 ----- 11
Db 3945 ILSGLAEGHVESLQSPDVMISQVNEPLVPSHSCTEGKILIESANPLENAGDSAVR 4004
QY 12 ----- 11
Db 4005 IECKSLRPLALEKQVILKEHSDNVVMPDQIISKEPVAIKAVQEOGRDLKE 4064
QY 12 ----- 11
Db 4065 SILSGIPEEORLNLKIQICRALQAAVASEQGLPSEWMLNIEKVEAVANTQEPRIHC 4124
QY 12 ----- 11
Db 4125 MYLTSKASVTEEYTIIEEDVPQMANLKEELDALCALIIEEDILTASBPRIQOAKT 4184
QY 12 ----- 11
Db 4185 SLQEMDSFGSQKVEPTEPEVESKYLLISPEVSYENQSRVYLDATVTKVASAV 4244
QY 12 ----- 11
Db 4245 SDEKODESLAPSEKEESESSESTEEVATVKIQEABEGPIKEDGPMIHTPLVTVSEGD 4304
QY 12 ----- 11
Db 4305 IYHLTTSITNAKENVNWFENKLVPSDEKFKLQDQNTYTLVYDKVNTEDHOGYVCEALN 4364
QY 12 ----- 11
Db 4365 DSGKTATSAKLTVKRAAPVILKRIEPLVALGHLAFTCEIOSAPVRVRQWFKAGHEIY 4424
QY 12 ----- 11
Db 4425 ESDKCSIRSSKYSLSLEILFTQVDDCGEYCKASNEYSVCTATLTVEAYPTFLSRP 4484
QY 12 ----- 11
Db 4485 KSLTFVGRKAAFLICVYTGPIETIMQKDGAALSPSPMKISDAEKHLELSNLTIOD 4544
Db 4545 RGVYSCASAKKFGADICQAEILIIDRPHFIKLEPVQSAINKRVHLEQYVEDRKVTVTW 4604
QY 12 ----- 11
Db 4605 SKDGOQLPPEKDYKICHEKIAIETLPLAKIKDSGTYVCTASNEAGSSSCSATVTVREPP 4664

QY 12 ----- 11
Db 4665 SFVKKVDPSYMLPGESARLHCKLKSPYIQVTFANKKELBSNTYKTFVNSEALIDI 4724
QY 12 ----- 11
Db 4725 TDVKEVDSGSYSCEAVNDVSDSCSTEIVAKPEPSTKTLLEPADIVGTNALLQCFVSGT 4784
QY 12 ----- 11
Db 4785 GPFEISFNDKKOIRSKKTRLSQKSLVCLCEIFSNADVGEYECVAVNEVKGCGMAT 4844
QY 12 ----- 11
Db 4845 HLLKEPTFYKXVDDILALGQTVTLQAAVRGSEPISTYTKMGQEVIREDKIKMSFSNG 4904
QY 12 ----- 11
Db 4905 VAVLIIPDVOISPGRYTCLAENAGSOTISGELIVKEPAKILIERALIQVTAGDPATLE 4964
QY 12 ----- 20
Db 4965 YTVAGTPELAKPKRYKQGRPLVASKKTRISFKNNVADLQKTSALHDSGYTTEISNEVGS 5024
QY 21 ----- 20
Db 5025 SSCETFTVLDRIAPFTFKPLKNDVSVYNGTCRLQCKTAGSLPMRVSPFKDKEIASD 5084
QY 21 ----- 20
Db 5085 KYRIAEVGTASLEITRVDMNAGNFTCRATNSVSGSKDSGALIVQEPSPVTKPKSKDV 5144
QY 21 ----- 20
Db 5145 LPSAVCLKSTNGSTPLTIRFKGNKELYSGGSCITTEALESLELYTKTSDSGTYT 5204
QY 21 ----- 20
Db 5205 CKVSNVAGVECSANLFEKPAFVEKLEPSOLKKGATOLACKVGTGPIKITWPAKD 5264
QY 21 ----- 20
Db 5265 REIKESKRAKMSVESTANVLRLLDVGIEDSGEYCEAONEAGSDHCSSIVYKESPYFTK 5324
QY 21 ----- 20
Db 5325 EFNPIEVLKEDVMLLAEVAGTPPEITWFKDNTILRSRKRYKTFIODHVSQILKEVA 5384
QY 21 ----- 20
Db 5385 ADAGEYQCHVTNEVSSISCSARVTLREPPSTFKIESTSLRGTAFAQATLKSLPTIV 5444
QY 21 ----- 20
Db 5445 TWLKDSELTEDDNIRKTFENNVASLYLSGIEVKHDKGYCOAKNDAGIQCSALLSVKE 5504
QY 21 ----- 20
Db 5505 PATIEEAVSIDVTOGDPATLOVFKFSGTKEITAKFKDQDELTLGSKYKISTVDTVSLK 5564
QY 21 ----- 20
Db 5565 IISTEKOSGEYTFEYQNDVGRSCKARINVDLLIIPSPFKIKLAKMDSINKSFIDECI 5624
Db 5625 VAGSHPISIQMKDQOELISASEKRYKFSPHDNTAFLEISQLBGTDSGYTCSAINKAGHNO 5684
QY 21 ----- 20
Db 5685 CSGHLVKEPPYFVEKPSQOVNPNTRVOLKALVGTAPNTIKMFKDKKELHSGAARSVM 5744
QY 21 ----- 20

```

Db 5745 KDDSTSLFLFAKKAIDSGTYICQLSDNDVGTATSKATLFVKEPPQFIKKRPPVLVLRNG 5804
QY 21 ----- 20
Db 5805 STTEPCQITGCPKIRVSMYLDGNETTAIOKHGISFLDGLATFOISGARVENSQTYCCAR 5864
QY 21 ----- 20
Db 5865 NDAGTASCSIELKVEKPEPTTIRELKVEVVKYSDVLECEVTGTPEEVTWLNKNEIRS 5924
QY 21 ----- 20
Db 5925 SKKTYLNDRVSVFNLHTKCDPSDTEGYQCIIVSNEGSGCSTRVALKEPPSFIKIET 5984
QY 21 ----- 20
Db 5985 TTVLKSATQSVIAGSPISITWLKDDQIILDDDNVYISFVDSVATLQIRSVDNCHSGR 6044
QY 21 ----- 20
Db 6045 YTCQAKNESGVERCFAFLVQEPQIYERAKSVDTREKDPNTECEVAGTEPELKVKTKLD 6104
QY 21 ----- 20
Db 6105 GKQIAPSRFSPENNVAAPRIOSVAKKODSGQYTFRENDPGSSCDATLRLVDQNIIP 6164
QY 21 ----- 20
Db 6165 SPTKRLTMDXVLGSSIHMECKVSGSLPISAQMFKDKETISAKTYLVCHERSVSLVN 6224
QY 21 ----- 20
Db 6225 NLELEDANTYCKVSNVAGDAGSGILTYKEPSPFLVKGQAIPDSIYEFKALKEGP 6284
QY 21 ----- 20
Db 6285 PFKIKWFDVLYSGPCPIGLEGSTPLNLYSVDSKGTQYCNVTDVDSGSCFTML 6344
QY 21 ----- 20
Db 6345 LVTEPPKVKKLEASKLYKACDSSRLCKIAGSPERIVFRNHELPAUSDKIWNTIDS 6404
QY 21 ----- 20
Db 6405 VAVIOMNLSTEDSGDFICEAONPAGSTSCSTKYIVKEPVSFPPIVETLKNAEVSLE 6464
QY 21 ----- 20
Db 6465 CELSGTPPEYVWKDKQOLNSKRYKIASKNFHTSHILNVDISDIGEYCKAONEVGS 6524
QY 21 ----- 20
Db 6525 DTCVCTVLEKPPRFRVSKLNSLVVAGAPALQASIRGAQPIFYQWLKEKEVIRESENI 6584
QY 21 ----- 20
Db 6585 RITEVENATLQFAKAEPAANAGKYICQIKNDGMEENATILVLEPAVIVKAGPMYTV 6644
QY 21 ----- 20
Db 6645 GETCTLECKVAGTPPLSVENWKDGLKLLTSQKHKFSFYNNKISSILILSVERQDAGTTFQ 6704
QY 21 ----- 20
Db 6705 VQNNYKSSCTAVVVDSDRAVPPSFTRLKNTGVLGASCTLECKVAGSSPISVAMFEK 6764
QY 21 ----- 20
Db 6765 TFIYSGAKQOTFSDNCTQLNLSIDSSDMGYTCVAAVAVAGSDECBRAVLIVQEPSPFVK 6824
QY 21 ----- 20

Db 6825 EEPLEVLPGKNVPTSVIRGTPEPKVWFRGARELVKGRDNVYFEDPTVALELFNIDI 6884
QY 21 ----- 20
Db 6885 SOSGETCVSNNAQASCTTRLFVKEPAFLKRLSDSVBPGSILIESTYTGLPSV 6944
QY 21 ----- 20
Db 6945 TWKDNFTTSKCNIVTTEKTCILEILNSTRKRDAGQYSCSEIENAGRDVCGALVSTLE 7004
QY 21 ----- 20
Db 7005 PPFVTELEPLAAVGDVSLQOQVAGPELTVSNTYKGTDLRPTPEXRTIYFTNNVATLV 7064
QY 21 ----- 20
Db 7065 FAKVINDSGETCKAENSIGTASSKTYFRIOERQLPSPARQLDIBOTVGLPYTLGR 7124
QY 21 ----- 20
Db 7125 LNSAPIQVCTRDGVLEDDENLQTSFVDNVATLKITQDLSHSGQYSCSASNPLGTAS 7184
QY 21 ----- 20
Db 7185 SSARLTAREPKSPPEIDIKPVSIDVITAGESADFECHVTGAQPMRTWSKDKELRPGNY 7244
QY 21 ----- 20
Db 7245 TITCVNTPLHLILKVGKDSQYTCQATNDYKDKMSAQSVKEPDKVKKLEASKVAK 7304
QY 21 ----- 20
Db 7305 QGESIQLECKISGPEIKVSNFRNDELHESKYNMSPINSVALLTINEASAGSDYIC 7364
QY 21 ----- 20
Db 7365 EAHNGVGDASCSTALTAKAPPVFTQKPPVYALKGSDVILQCEISGTPPEYVWKDKRQ 7424
QY 21 ----- 20
Db 7425 VRNSKKFKITSKHFDTSLHILNLEASDVGEYHCKATNEVSDPCSVKFKRPPVKL 7484
QY 21 ----- 20
Db 7485 SDSTLIGDAVELRALIVEGFOPISVWMLDKGEVIRESENTRIISFLDNTATLQLSPEAS 7544
QY 21 ----- 20
Db 7545 NSGRYICQIKNDAGMECSAVLVEPARIIEKPEPMVTITGNPFALRCVYTGTELSAK 7604
QY 21 ----- 20
Db 7605 WFDGRLEASDKHITFLINKVASLKIPEAMSDKGLSYFEKNSVGNCTVSVHSDR 7664
QY 21 ----- 20
Db 7665 TVPPSFTRLKLDVNAILGASVLECRVSGSAPISVGWFOGDNEIYSGPCOSSFSENVCT 7724
QY 21 ----- 20
Db 7725 LNSLEPDTGITYTCVAAVAVAGSDECSAVLIVQEPSPFQRPDSVEVLPGMSLFTSVI 7784
QY 21 ----- 20
Db 7785 RGPPEKVKWFKSRELVPESCNISLEDFVLELFEVQPLESGDYSCLYTNDAGSASC 7844
QY 21 ----- 20
Db 7845 TTHLFVKEPAFVKKRLADSVETGSPIVLEATITGTTPJISVSHIKDEYLLISQERCSITH 7904
QY 21 ----- 20
Db 7905 TEKSTILELESTIEDYAYQYSCLIENAGODICEALVSVLEPPYFTEPLEHVAIVGEP 7964

QY	21	-----	20
Db	7965	TLQCKVDGTEIRISMYKEHTKLNSAPAYKQFNNAVSLINKVDHSDVGEYCKADNS	8024
QY	21	-----	20
Db	8025	VGAVASAALVYKERKLPPEPARKLDVHETLGPVAFECRINGSERLOYMTKDVLLK	8084
QY	21	-----	20
Db	8085	DDANLQTSFYHNVALIQLQTDQSHIGQVNCASNPLGTASSAKLLISEHEVPPFPDLK	8144
QY	21	-----	20
Db	8145	PVSVDIALGSGTFCCHVTGTAPIKIYKANDNEIRPGNFKMTLVENTATLVLKVGK	8204
QY	21	-----	20
Db	8205	DAGQYCYASNINAGKSCSAHLGVOEPPREIKTLEPSRIYKODEFTRECKIGGSPEIRY	8264
QY	21	-----	20
Db	8265	LWTKDETEIQESSKFRMSFVDSVAVLEMHNLSEYDSDYTCEAHNAGSASSISLKYE	8324
QY	21	-----	20
Db	8325	PPIFRKKPHDIEFLKADVHLECELOGTPEFNVSWKDKRELNGKRYKIMSHPITSIH	8384
QY	21	-----	20
Db	8385	ILNVDAADIGEYQCKATNDVGSJTCVGSIALKAPPEPVKLSDISTVQKEVOLQTTIEG	8444
QY	21	-----	20
Db	8445	AEDISVWFADKEIYRESNDIWSISEMIATLOFSRVEPANAGKYTQIKNDAQOECF	8504
QY	21	-----	20
Db	8505	ATISVLEPATIVKPEISIVTGTDTLCEQVAGPELSTKFKKDKELTSDMKYKISFF	8564
QY	21	-----	20
Db	8565	NKVSGLKIINVAAPSDGSYTSFEQNPVGKDSCTASLOVSDRTVPSTFKKLETNGLSGS	8624
QY	21	-----	20
Db	8625	SVVMECKVYGSPPISVSMFHGNEISSGRKQYOTLTIDNTCALTNMLESDSGDYTCIAI	8684
QY	21	-----	20
Db	8685	NMAGSDECSAPLTVREPPSFVQKPDMDVLTGTNVFTSIVKGTPEFSVMEKSGSEIWP	8744
QY	21	-----	20
Db	8745	GDRCNVLEDSVAELELFDVDTSGSEYTCIVSNAGKASCTHLYIKAPAFYKRLNDY	8804
QY	21	-----	20
Db	8805	SLEKGPILLEGFTGTGPPISVTWKNGINWTPSORCNITTEKSAILEIPBSYVEDAQ	8864
QY	21	-----	20
Db	8865	YNCYIENASGKSCSAQILLIEPPYFKOLEPVKVSGDSASLOQLAGTPEIGVSWYK	8924
QY	21	-----	20
Db	8925	DYKLRPTTYKXHFNNNAVILFVNOVINDSGEYICKAENSVGEVASASTFLVQXQLBP	8984
QY	21	-----	20
Db	8985	SFSROLNDVQETVGLPVVFDCAISGSEPISVSWYDGRPLKDSFVQTSFLDNTATLNI	9044
QY	21	-----	20
Db	9045	KIDRSIAGYSCATNPISGASSARLITTEKNPPEFDIRLAPDAVYGESADECHVT	9104
QY	21	-----	20
Db	9105	GTQPIKVSMAKDSREIRSGKQYQISYLENSAHLVLYKVDKDSGQYTCYAVNEVKDCT	9164
QY	21	-----	20
Db	9165	QOLNIKERLIPPSFTKRLSEIVEETEGNSFKLEGVASOPITVAMYKNNIEIOFTSNC	9224
QY	21	-----	20
Db	9225	ITFKNNTLVIQYRKAGMNDAGLYTCQVSDAGSALCTSSIVYKEPKPVPFDQHLPTV	9284
QY	21	-----	20
Db	9285	SSEYVOLSCHVQSGEPRIQWLKAGRELKPSDRCSFASGTAVLELDVAKADSGDY	9344
QY	21	-----	20
Db	9345	CAASVAGSDTTKSVTLKDPAPAPATKKAVDGRLEFVSEPOSIRVEKTTATFLAKV	9404
QY	21	-----	20
Db	9405	GDDPIPVKWKFGKROLNGSGRVEIHOKGDEAKLEIDTTRTDSGLYRCVAFNEHGEIE	9464
QY	21	-----	20
Db	9465	SNVNLQVDERKKOEKIEGDLBAMLEKTPILKKGAEIEEIDIMELKNVDREYKAYAM	9524
QY	21	-----	20
Db	9525	YQITPFRGLQAFELKQSOEETHRLTEIEIENSRDEKEEELVSFTQRLSOTEPVT	9584
QY	21	-----	20
Db	9585	LKIDENQVLKNDVAPEIDIKINYPEIKLSWYKGTPELPSDKFELSDGDMHTLRVK	9644
QY	21	-----	20
Db	9645	NCOLKQGNTRYLCGPHIASAKLVIEPAMERHLQDVLKSGQCTQTCQFSVNVKSEW	9704
QY	21	-----	20
Db	9705	FRNGRIILPQGRHKTEVBHAKYHKLTIADVRAHDQGYTCKTBDLETSABLRIEAPTOFT	9764
QY	21	-----	20
Db	9765	KRIQIVVSEHQSATFECEVSFDALVTWYKGPTELSEOSQYNFRNDGRCHYMTIHNVTP	9824
QY	21	-----	20
Db	9825	DDEGYSTIARLEPRGENRSTAEIYLTKKIKELKEPDIPIPSRVPITPPIRAVPPEEI	9884
QY	21	-----	20
Db	9885	PPVAPPIPLLPPEEKPPPKRIEVTYKKAVKKAKKVAKREMTPREIYKPPPPPT	9944
QY	21	-----	20
Db	9945	TLIPAKPEIIDVSKAEVKNITITRKKEVQKEKFAVYKQAVHKEKRVLESPEPY	10004
QY	21	-----	20
Db	10005	DELEVEPYTEFBQPYTHEPDEDEYEEIKVAKKVHEHEDEDEGOEYTEREGYDEGE	10064
QY	21	-----	20
Db	10065	EEWEAVQEREVIOQKEVYEEESHERRKVPKVPKAPPPPVYIKKPVIEKIKTSRRME	10124
QY	21	-----	20

Db 10125 EKKVQTKVPEVSKKIVPQKPSRTVQSEVIEVKVPAVHTKKAVTISEKGFASHTSEBV 10184
QY 21 ----- 20
Db 10185 SVTVPEVOKIIVLEEKIHVAVSKRVEPPKVPBELPEKPAEVAAPVPIPKKVEPPAKVP 10244
QY 21 ----- 20
Db 10245 EYPKKVPPEKKVPVPKKEBPAPKVPBPVKPVEBEKIIPVPAKKKEAPPAKVPVQK 10304
QY 21 ----- 20
Db 10305 RVVTEEKITIVQRESPPAVPEIPKKKVEBEKKVPKRESEVPPKVPALPKKVPDE 10364
QY 21 ----- 20
Db 10365 EKVAVVPVPAKKAPPAPPAVSKTVVEERFVAEKLSPVQKVEVTHSEVSAEEMSY 10424
QY 21 ----- 20
Db 10425 SEEBEGSISVREBEREEBEAEVTEYEVMEBEPEYVEBEKLIISKRYEAPAEVTER 10484
QY 21 ----- 20
Db 10485 QEKKIIVLKPPIPAKIEBPPPAKVPAPKIVPEKKVPAPVKKKVPVPKVPBEPKVPV 10544
QY 21 ----- 20
Db 10545 EKKVPKVIKMEBPLPAKVTERRMOITQBEKLVAVTKKEAPKAVPPEPPRAVDEKV 10604
QY 21 ----- 20
Db 10605 LKLPKREBPPAKVTEFRKRVVKEEKVSTLAPKREBPQIKVYIMBEKRAVTLSEAV 10664
QY 21 ----- 20
Db 10665 SVQREBEYEEYEEYKKEPEYEPTEYDQYBEYERNEYRYEBHEEYITPEKPIPVK 10724
QY 21 ----- 20
Db 10725 VPBPVTKRPAKPAKVLKAVBEKVPVPIPKKIKRPPKVPBEKKVVEBKIRISTIK 10784
QY 21 ----- 20
Db 10785 REKQVTEPAKVPKPKRVAAEKVPVPRKEVAPVPRVPEVKLEPEBEVAFEEEVYTH 10844
QY 21 ----- 20
Db 10845 VEELVVEEBEYIHEEBEFTTEEVVPIPVKVPBPVKPVEBKVPVPPKKEAPPAK 10904
QY 21 ----- 20
Db 10905 VPBPVKPPEKVPVLIKKKEPPPAKVPBPVKVPBEKVPVPPKVEAPPAKVPBPVK 10964
QY 21 ----- 20
Db 10965 KPVPEKIVPAPAPKVEAPPAKVPBPVKLIPEBKPTVPKPKVEAPPKVPKREBPV 11024
QY 21 ----- 20
Db 11025 PVALPQEBEVLFEELIVPEBEVLPDEBEVLPDEBEVLPDEBEVLPDEBEVLPDE 11084
QY 21 ----- 20
Db 11085 EBYVPEEBEYVPEBEVLPBPVKPVPAPVPEIKKVTETKVVIPKKEAPPAKVPBPVK 11144
QY 21 ----- 20
Db 11145 KVEBKRIILPEBEVLPVEVTEPEEBPISSEIPEEPPSIEVEEVAAPVPEVINKAV 11204
QY 21 ----- 20

Db 11205 PEADTPVKKVAPPAVSKKIPBEKVVPVQKKEAPPAVPEVKKVLPKKEA 11264
QY 21 ----- 20
Db 11265 VPBAKGRIVLEKVSVAERQVYVYKERTLELVVAEVEIPEEBEFHEVEFEESEFHE 11324
QY 21 ----- 20
Db 11325 VEEFKLEQHNEBHEHVEKVRVLEVEFAEVEVFEKPAKPEKPEISKIIPEKPKPT 11384
QY 21 ----- 20
Db 11385 KVPKREPPAVPEYKPIVVEKVRVDEBRVPPIKVPDVLPEKVPVPEKVPVPAK 11444
QY 21 ----- 20
Db 11445 PEAPPPVPEAPKEVVPPEKKVPBPVKPEVPTKVPVPAVPAVEKKVPEALPKPESP 11504
QY 21 ----- 20
Db 11505 PPEVPAKREVPPEKKVPAPPKPEVTPVKVPAPKEVVPPEKKVPVPPKPEVPPTKV 11564
QY 21 ----- 20
Db 11565 PEVPAVAPPEKKVPBATIPPKPSPPPEVFEPEBEVALPEPPAIVEPEBAPAPVTPP 11624
QY 21 ----- 20
Db 11625 KKPVEKAPAVVAKKPELPLPVKVPEVPKEVVPPEKKVPLVPPKKEAPPAKVPBPVK 11684
QY 21 ----- 20
Db 11685 PEKVAVPRKKEVPAPKVPPEKVPLEKPAVPEBAPESPPPEVTEPEPEIAPBEEIA 11744
QY 21 ----- 20
Db 11745 PEEKVPVVAEERPEVPPPAVPEBKIIPEKKVPVTKKPAAPPKPEPEKVIKPKL 11804
QY 21 ----- 20
Db 11805 KRPVPPPAPEKEDVKEIFQIKAIIPKKVPEKQVPEKVELPLKVGGEKKVRKLLP 11864
QY 21 ----- 20
Db 11865 EKKPEKEEVLKSVLRRKPEEBEPKPEKLEKVAVPEPPPKVPEEVPVPTKR 11924
QY 21 ----- 20
Db 11925 EKKIPEPTVPBIKPAIPLPAPPEKPKPEAKVTIKPPVEDEPTPIAPVTVPVGKA 11984
QY 21 ----- 20
Db 11985 EAKAPKEBAKPKGPIKGVPKTSPPIAEERKILPBGSGGEPDEAPFTYQLAAPLKF 12044
QY 21 ----- 20
Db 12045 VAKIKDIIITSEFVGSNAIFCLVSPSTAITTKKOSNIRESPKHPFIADGDKRIHI 12104
QY 21 ----- 20
Db 12105 IDVQSLDAGETCVLRIGKKEKSTAKLVBEELPVRFVKTLEEVYVKGQPLYLSCIA 12164
QY 21 ----- 20
Db 12165 KERDVWRKGRIVVEKPGRIVPGVIGIRALFTINDADDTAGITVVENANNIECSC 12224
QY 21 ----- 20
Db 12225 VAYVEVINDVLKPIHQHVAKKGTAIIRACDIATPNIKWKGVDEIAPBPNOKTEILR 12284
QY 21 ----- 20
Db 12285 DGNHLYKIKNNAPEDIAEYAVEISGRVPAKLTJGERREVELKPIEDVTIYKESASFD 12344


```

QY 21 ----- 20
Db 12345 AEISADIPGQWKLKGBELLAPSPICEIAGEGKREPLIRKVKLDQAGEVLYQALNATTA 12404
QY 21 ----- 20
Db 12405 ILTVKEIELDFRVLKDVTPERRORARPECVLTREANVYMSKGPDIKSDKFDIADGK 12464
QY 21 ----- 20
Db 12465 KHLIYNSQDFDEGVYTAEEGKTSARLFVIGIRLKEKSPLEDQYKEGETATFVCEL 12524
QY 21 ----- 20
Db 12525 SHEKHWVWFKNDAKLHTSRTVLISBEGKTHLEKEVTLDDISOIKQVKELSSTAOLK 12584
QY 21 ----- 20
Db 12585 VLEADPTTVKLHDKTAVEKEDEITLKEVSKDVPVKWFKDGEIIVPSPKYSIKADGLNRI 12644
QY 21 ----- 20
Db 12645 LKIKKADLKDGEYVCDGTDKANTVEARLIKVEKPLVGEVYGETAHEIELESEP 12704
QY 21 ----- 20
Db 12705 DVHGWMLKGOPLASPDCIIEDEKRRHILANCOIGMTGEVSFOANAKSAANLKVKE 12764
QY 21 ----- 20
Db 12765 LPLIFITPLSDVVKVEKDEAKFECEVSREPTFRMLKGTQETIGDDDELKDGTHSNV 12824
QY 21 ----- 20
Db 12825 IKSAAFEDEAKYFEADEKHTSKLIIEGIRLKLPLTPLKDVTAKEKSAVTVELSHDN 12884
QY 21 ----- 20
Db 12885 RVKWFKNQDRIHTRSVSMODEGTHSTIYFKDLSIDDTSOIRVEAMKSSBAKLTVLEGD 12944
QY 21 ----- 20
Db 12945 PYFTGLQDVTGVEKDEVILOCEISKADAPVKWFKDGEIKPSNNNAVIKADGKRMLILK 13004
QY 21 ----- 20
Db 13005 KALKSDIGQYTCDCGTGTSKGLDIEDREIKLVPLHSEVMEETARFETISEDDIHA 13064
QY 21 ----- 20
Db 13065 NWKIKGALLQTPDCEIKEGKTHSLVLHNCRLDQGTGVFOANAKSSAHLNKPRAVIG 13124
QY 21 ----- 20
Db 13125 LMLRLKDVYTAGETAFDCDELSEYDIPVEMYLKGGKLEPSDKVVPRESKGVHTLILROY 13184
QY 21 ----- 20
Db 13185 KLEADAGEVOLTANDKTHANLFWKEPVEFTKLEDOVTEGATVALECEVSRENKAVK 13244
QY 21 ----- 20
Db 13245 FKNGTILKSKKYEIVADGRVKLVHDCIPEDIKITTYCDAKDFTKSCNLNVPPHVEH 13304
QY 21 ----- 20
Db 13305 RPLVDLQVREKEMARECELSRENKAVKPKDGAELKGGKTYDIISKGAVALIYINKCLL 13364
QY 21 ----- 20
Db 13365 DDEAETSCETVARTSGMLTVLEEAENVTKNLNANIEVSETDTIKLCEVSKPGAEVIMWK 13424

```

```

QY 21 ----- 20
Db 13425 GDESIIFTGVEIITBGRKRILYQNAHLDEDAQNYNCRILPSSRTDQKRVYHELAEFISK 13484
QY 21 ----- 20
Db 13485 PONLEIEGKAEFVCSISKESFPVQKRRDKTLESQDKYVLAQDKRRLVYKNTATLQD 13544
QY 21 ----- 20
Db 13545 MGYVYVWGAARAABHVTIEKLRIYVPLKDTYKEDQEVFNCEYNTBGAKWFRNDE 13604
QY 21 ----- 20
Db 13605 AIFDSSKYLIILOKDVIYTLIRDAHLDDQANTVNSLTNRHGENVKSANLIVEEEDLRIV 13664
QY 21 ----- 20
Db 13665 EPLKDIETMEKKSVTFWCKVNLNVTLKWKNGEYVFDNRVSYRVDYKXHMLTIKDGCF 13724
QY 21 ----- 20
Db 13725 PDGERTITRAGQDKVAELLITAPTEVEHELEDQVTEFEDAVFSCQLSRREKAVKMYR 13784
QY 21 ----- 20
Db 13785 NGREIKGKKYKPKENDGSIHRLIKDCRLDDECEYACGEYEDKSRARLFVEELIPEVILIRP 13844
QY 21 ----- 20
Db 13845 PDLILPAGADVFLAELNKKVEVQWMLNNNVVQGRHOMKSEKIHRLQICDILKPRD 13904
QY 21 ----- 20
Db 13905 QGRIYRIANDKEARALELAAPKIKTAQDQDVYVGNPLVWVPYDAVPAKAEWFKEN 13964
QY 21 ----- 20
Db 13965 EPLSFTITDTAEQVFRILKAKKGDKGRKIVLONKHGKAGETINLKVIVPGVRNLE 14024
QY 21 ----- 20
Db 14025 VMTFDEGVSLAWEPLDQSGSKIGYVVERRDIKRKTWVLATDRASECEFTVGLORGG 14084
QY 21 ----- 20
Db 14085 VETLFRVSARNRYGEPVETDNVBARSKYDVPBPPLNVTITDVNKRGVSLTMEPPERD 14144
QY 21 ----- 20
Db 14145 GGAETINVYIELRDKTSIRMDTAMTVRABDLSATVTDVVEGOEYSPRVAQRIGVGRPS 14204
QY 21 ----- 20
Db 14205 AATPEYKADPTLERSPPVNLTSSDQTOSSVOLKWEPLKDGSPILGYIIERCEBGRDN 14264
QY 21 ----- 20
Db 14265 WIRGNKLEVPELTYKVLGKGNKTYLRYSAENKAGYSDPSIILGPLTADDAFVBPYIDL 14324
QY 21 ----- 20
Db 14325 SAFKQGLEIVYENPILIVPSTGYPRPATWCFQKULETGDVRVKKKTLISANAEIVSPS 14384
QY 21 ----- 20
Db 14385 ERSQKITYLAKENKVTISGEIDVNVIAKRSAPKELFGDITKDSVHLTMEPPDDGGS 14444
QY 21 ----- 20
Db 14445 PLTGVYVERREVSRTWTKVMDFVTDLEETVBDLVQGEYLFKVCARNNKCGSPAYYDE 14504
QY 21 ----- 20

```

DD 14505 PVMKSTPATVPOBPENKMDRTANSIFILMDPKNDGSRIGYIVERCPHSGDKWVC 14564
 QY 21 ----- 20
 DD 14565 GEPVAETKMEVTEGLEBGRATAYVKALNRGASKPSRTEIQAVDTQEAPEILPDKLL 14624
 QY 21 ----- 20
 DD 14625 AGLTWAGTKIELPATVTCAPPEKITWTAKMDLKKODKRIITENVRKSTVITVDSKRD 14684
 QY 21 ----- 20
 DD 14685 TGTITIAVAVCGRATAVEENVLDKGPAPADITDVTNESCLLWNPBRDGSKITN 14744
 QY 21 ----- 20
 DD 14745 YVERRATDSEVHKLSIVKDTNFKATKLIPIKKEYIPVAAEMNTGVCEPVQASPTAK 14804
 QY 21 ----- 20
 DD 14805 YQDPGCPPTLBPESDITKAAVILTWCEPDGSGSITGTVTERLDPTDKWVRCKMAY 14864
 QY 21 ----- 20
 DD 14865 KOTTVYKGLNKKKYRFRVLAENLAGPKPSKSTEBILIKDIDPMPGKPTVAVGK 14924
 QY 21 ----- 20
 DD 14925 TSVRLMNTKPEHOGANIESVIEMKTGTDENVRVAESVPTQHLPLGLMEQOESFRV 14984
 QY 21 ----- 20
 DD 14985 RAVKAGESESESBDVLCKREKLYPSPRMLEVINTKTADLKWTPEKXGSGPITN 15044
 QY 21 ----- 20
 DD 15045 YIVERKDYRRKQMTVDITWKOTKQVYTPLEGSLYVRVAENAIQSDYTEIEDSVLA 15104
 QY 21 ----- 20
 DD 15105 KOTTTGPPYALAVVDYTKRHVDLKWEPKNDGGRYQRYVEKKERLGRNWKAKTA 15164
 QY 21 ----- 20
 DD 15165 GPDCNFRVTVIEGTEVOFVRAENAGVGHSEPTBILSIEDPTSPSPDLCHVTAG 15224
 QY 21 ----- 20
 DD 15225 RKHIAIAWKPEKNGSPIGYHVMCPVCTEKWVRVNSRINDLKFRVEGVVPDEYV 15284
 QY 21 ----- 20
 DD 15285 LRVAAVNAIGVSESEISENVAAKDPCKPTIDLETHDIIIVIESEKLSIPVPRVAVPT 15344
 QY 21 ----- 20
 DD 15345 VSMHDKREVRASDRLTKMKNHISAHLEVPRSVRADAGIYITLENKLGASATASINVKI 15404
 QY 21 ----- 20
 DD 15405 GLPGCKDIKASDITTKSSCKLTWEPPEPDGCTPILHVLVLENRAGRTIYIVSGENKLS 15464
 QY 21 ----- 20
 DD 15465 WTVADLIINGEIEFRVAVNVVGGGEYIELKNVIAODPKPPDPYDVEVHNPTAEAMT 15524
 QY 21 ----- 20
 DD 15525 ITWKRPDYDGSKIMGYILIERIANGEBMKRCNHLVPLIYTAKGLEGEKETOFRVRAE 15584
 QY 21 ----- 20

DD 15585 NAAIGSEPSRATPPTKAVDPIADPFVILMTSLFVKRGDELALDASISGSPYITTWIKDE 15644
 QY 21 ----- 20
 DD 15645 NVIPEEIKKRAPIVRRKCEVOEEEPVLPJTORUSIDNSKSGESOLRVRSRLRDPHG 15704
 QY 21 ----- 20
 DD 15705 LYMIKENDHGIKARCTVSVLDPGPPINVFEDIRKTSVLCWKEPPLDDGGEIINTY 15764
 QY 21 ----- 20
 DD 15765 LEKKDKTPDSEMIYVTSILBCKISYTKLIEGKEYLEVRBAENRFGCPGVSKPLVAK 15824
 QY 21 ----- 20
 DD 15825 DPGPPDAPDKPIYVDVYINSMLVKNRPKXGSPILGYMLEKREYNSTHMSRVNKSILN 15884
 QY 21 ----- 20
 DD 15885 ALKANVDGLLEGLYVFRVCAENAGPKFSPSPDKTAHDPIBPGPIRVTIDTSSTY 15944
 QY 21 ----- 20
 DD 15945 IELEPPAFNGGGEIVGYVDKOLVGTNWSKCTEKMIKVRQYTVKEIRGADTKLRVS 16004
 QY 21 ----- 20
 DD 16005 AVNAAGCPPGTQPVYTAEPQEPRAVELDVSVKGCIQHMAGTKLIRIPAVYGRPVTKV 16064
 QY 21 ----- 20
 DD 16065 WTKKEGELDKDRVVIDNVGKSELIIKDALKRDRGYVITATNCGSRFAAAAVEVDVP 16124
 QY 21 ----- 20
 DD 16125 GPVLDKPYVYTRKCKLWMSDPEDDGSEITGFIIEKDAKMHTRQPIETESKCDIT 16184
 QY 21 ----- 20
 DD 16185 GLEGEYKFRVIAKNNKFCGPPVEIGPLAVDPLGPTSPRLTYTERKSTIILDKKE 16244
 QY 21 ----- 20
 DD 16245 PRSNGSPYQGYIIEKRKDKDPFERVNRKLCPTTSFLVNIIDEHONYEPRVAVNEIGE 16304
 QY 21 ----- 20
 DD 16305 SEPSLEPLNVYIODDEVPTIKRLSVRGDTIKVXAGEPVHIDADYIGLPMKXIEMSKNET 16364
 QY 21 ----- 20
 DD 16365 VIEKPTDALQITKEEYNSAEKTELISIPRAVREDQTYVVTASNRLGVSFRVAVEYDR 16424
 QY 21 ----- 20
 DD 16425 PSPPRMLAVTDIAKESCYLITMAPLDNGSELTHYVIDKRDASRKAKAEVYTNVAKER 16484
 QY 21 ----- 20
 DD 16485 YGIWKJIPNGOZEFRVAVNKKYISDECKSDKVIYODPYRLGPPGKPRVLARTKSMUY 16544
 QY 21 ----- 20
 DD 16545 SMTPLDNGSPITGYMLEKREGBPYMSVSRAPITTKVGLGVEFNPRLLEGVYQFR 16604
 QY 21 ----- 20
 DD 16605 AVALNAGIGPSESESDPEVADPIRPPGPPSCPEYKOTKSSISLGKTPRAKDGSPK 16664
 QY 21 ----- 20
 DD 16665 GYIVMOEBCTTDMKRVNEDKLIITTCGVPNLKEIKRYRFRVAVNAEAGESESPDITG 16724

QY 21 ----- 20
Db 16755 EIPATDIEEVEFIDIGAODCLVCKAGSQIRIPAVINGRPPKRSWFDGKAKKAMDG 16784
QY 21 ----- 20
Db 16785 VHDIPEDAOLEAENSAYIIPECKRSHTGKYSITAKKAGQATANCNVKADVPGPBKD 16844
QY 21 ----- 20
Db 16845 LKVSDFRSGRLSKWKMDDDGDRKIGYIEKRTIDGKANTKYNPOGSIFFVPPDLS 16904
QY 21 ----- 20
Db 16905 EQQFFRVRAENRPGIGPPVETIQRTARDPIPPDPPIKLKGLITKNTVHLSMKPKPN 16964
QY 21 ----- 20
Db 16965 DGGSPVTHYIECLAMPDTGTRKEMRQCNKRDYBELQFTYEDLVEGGEYFRVKAANAA 17024
QY 21 ----- 20
Db 17025 GVSRRSATVGPDCQRPDMPSIDLKEMEYEGSTNVIYAKIKGVPPLTWFKAPPK 17084
QY 21 ----- 20
Db 17085 PDNKEPVLYDTHVKNLYVDTCIYIIPQSRSDTGLTITAVNNLGTASKEMRLNLVIGRP 17144
QY 21 ----- 20
Db 17145 GPPVGPILFEBSADQMTLSWFPKDDGSKITNYIEKREANKTWVHVSSEPKCTYT 17204
QY 21 ----- 20
Db 17205 IPKLLEGHEVFRIMAONKIGIGEPIDSEPEIARNLESVGADPKPTVSVTRNSMTVM 17264
QY 21 ----- 20
Db 17265 EEPYDGGSPVTVWLEMKDITSKRKRVNRDPIKANTLGVSITVIGLIGSDYQRPVYA 17324
QY 21 ----- 20
Db 17325 INAGVPASLPSDPATARDPIAPGPPKVTMDWTKSSADLWSPPLKDGSKVTCYIY 17384
QY 21 ----- 20
Db 17385 EYKEGEKEKMKGDKDEVRGTKLWVGLKEGAFYKFRVSAVNTAGIGEPGEVTVIEMKD 17444
QY 21 ----- 20
Db 17445 RLVSPLQLDASVDRIVHAGVIRIIVASGKPPPTVWNNERTLQEAETIETTAIS 17504
QY 21 ----- 20
Db 17505 SSMVINKQQRSHOGVYSLAKNAGEKRTIIVDVIDVPGVGIPLAHNLTFNSCKLW 17564
QY 21 ----- 20
Db 17565 FSPEDDGGSPITNVYIEKRESDRANTPVYITVRONALVQGLLOGKAFIRIAENSIG 17624
QY 21 ----- 20
Db 17625 MGFVETSEALVIREPITVERPEDELEKVEIKNTVTLWNPKNYDGGSELIINYVLESN 17684
QY 21 ----- 20
Db 17685 IGTEKREKVTNNILSRKITYVGLKESDYEVKVSANVIVGQKPSFCYKPIPCDELA 17744
QY 21 ----- 20
Db 17745 PTLHDFRDLTIRGEAFALIGRYSGRPKPKVSNPKDEADVLIEDRTHIKTTPTATLAE 17804

QY 21 ----- 20
Db 17805 KIRAKRSDSGKCYVYENSTSGRKGCQVNVVRPGPPVPSFDEVTDIYNTISMKPPL 17864
QY 21 ----- 20
Db 17865 DGGSKITNTYIERKEVKADWNPVTASAKTYCKVSKLLEGKDYIFRIHAENLYGISDP 17924
QY 21 ----- 20
Db 17925 LVSDSKAKDRFRVPDADOPYTEVTKDSALVTNNKPHDGKPIYNTYIERKETSRR 17984
QY 21 ----- 20
Db 17985 ARVTKDPIHETKFRVDPDLLEGQYERVAENEIGIDPSPSKVPAKDIPLAKSPV 18044
QY 21 ----- 20
Db 18045 NPEALDTCNSVDLWOPPRHDSKILGYIYEQVGDENRANHTPESCEYKRYVT 18104
QY 21 ----- 20
Db 18105 GLRDGQYKFRVLAVNAAGESDPAHYEPVLYKDLPEPELLIDANMANECHIKVGDTLR 18164
QY 21 ----- 20
Db 18165 LSAIINGVFPKVTWKEKEDAPTKARIDVTPVSKLEIRNAHDEGGIYSLTVENPAGS 18224
QY 21 ----- 20
Db 18225 KIYSAVVLVDKGPRLDVESEIRKDSYLTWKEHLDGGSVITNYVERRDVAQMS 18284
QY 21 ----- 20
Db 18285 PLSATSKKSHFAKHLNEGNYLFRVAENQYGRGFVETPRIRALDPLHPPGPKDLH 18344
QY 21 ----- 20
Db 18345 HVDVKTESLVNWKRPDRDGSPIGTLYEQEETQDIKFKVTWNLCECVTLQOGKT 18404
QY 21 ----- 20
Db 18405 YRFRVAENIVGLGPDITIPLEQEKLVPSVELDKLISLVVAKAGTVFPAIIRGV 18464
QY 21 ----- 20
Db 18465 PVPTAKTTDGSSEIKYDEHYIYEDNFSVYLTKNCLRDRGEGYQITYSNAAGSTVAH 18524
QY 21 ----- 20
Db 18525 LTVLDVPGPPTGINILDVTPRHMTISNOPPDGGSFVINYIYERKQTRKDTMGVSSG 18584
QY 21 ----- 20
Db 18585 SSKTKLKIPLQKGEYFRVAENKIGVPLDSTPIVAAKHTSPSPGKPVVTDITE 18644
QY 21 ----- 20
Db 18645 NATVSWTLPKSDGSPITGYMERREVTGKVRVNTPIADLKERVYGLXGNTYFRV 18704
QY 21 ----- 20
Db 18705 FRENLAGLSKSPSSDPIKACRPPIPPGPPINPKLDRSRETDLWTKPLSDGSPILG 18764
QY 21 ----- 20
Db 18765 YVVECKPQIAQMNRLINDELIRQCAFVPGULBSNEYFRITKANIVGEGEPBELAESV 18824
QY 21 ----- 20
Db 18825 IAKDILHPEVELDVTCHDVITVRVQGITRILARVGRPEPDITWTEGKVLVREKRVDL 18884
QY 21 ----- 20

```

Db 18865 IODLPVELOJKEAVRADHGXIIISAKNSSGHGSAIIVNLDPRGPCOLXATNTYTKEN 18944
QY 21 ----- 20
Db 18945 CTISMENPLDNGSEIYNFIEYERKPNQKMSIYASDVTRILKANILANNEYFRVCAE 19004
QY 21 ----- 20
Db 19005 NRQVGWPTIETKPIILAINPIDRGEPEENLHIDKCKTPVYLKMRPDYDGGSPNLSYH 19064
QY 21 ----- 20
Db 19065 ERRLKSGDMEVRNKKSIKETHYVADRCVENQIYEFRTQKNESGESDVKTEEVYKED 19124
QY 21 ----- 20
Db 19125 LQKPEVLDKLSGLTVKAGDTIRLEAGVRGKPEPEVAMTKDKATDILNSPRVKIDTRAD 19184
QY 21 ----- 20
Db 19185 SSKFSLTKAKRSDGKTYVATNTAGSFVAYATVYVLDKFGPRNKLIVDYSSDRCTYCN 19244
QY 21 ----- 20
Db 19245 DPEPDGGEIONTILEKCEKRWVSGYSATVLTGITYVRLIEGNEYIFRVAENKIG 19304
QY 21 ----- 20
Db 19305 TGPTESKPIVIAKTKYDKPGRDPPEVTKVSKEETVYVWNPPEYDGGKSTGYFLEKKEK 19364
QY 21 ----- 20
Db 19365 HSTRNVPVNSAIEBRKAKYONLIPDHEYOFRVAENIGESLSPRYVANDPIEP 19424
QY 21 ----- 20
Db 19425 GPTNFRVVDTKHSITLGMGRPYTDGAPITIGYVEERPRIADSPDEGRCNMAAOL 19484
QY 21 ----- 20
Db 19485 VRKEETVTLSDENOYEFRVCAONQVIGRAPALKEAIKREITIEPPEIDLDASMKLYI 19544
QY 21 ----- 20
Db 19545 VRAGCPILFALVNGRPAPKVTMRKVGIDNYRRGOVDLVDIMAFVLIPNSTRDSGKYS 19604
QY 21 ----- 20
Db 19605 LTLVNPAGEKAVFVNVRYLDPGPVSDLKVS DVTKTSCHSWAPPENDGSGQVTHYIYK 19664
QY 21 ----- 20
Db 19665 READRKTSTYTPVEYKTSFHYTNLVPSNEYFRVTAVNEYGVPVDPKPYLASDPLS 19724
QY 21 ----- 20
Db 19725 EBDPRKLEVTENKTSATLAMPPLRDGAKIDGYITSREBOPADRMTEVSUVKLS 19784
QY 21 ----- 20
Db 19785 LVVTGLAKGKKYKFRVARNVAVSLPREABGVTAKEQOLLPRKILMDEQITTAGKKLR 19844
QY 21 ----- 20
Db 19845 IEAHYVGRPHPTCKKKKGEDEVYTSHLAVHKAQSSSILIIKDVTRKDSGYYSLAENSS 19904
Db 19905 GTDTQIKRVVADAPGPOPEDISIDADACSLSMHTPLEDGSNTINYLVEKCDVSRG 19964
QY 21 ----- 20

```

```

Db 19965 DWYATALASTKTSCHVGKLLIPGOEYIFRVAENRGISEPLTSRMAQPFPGVPSSEKN 20024
QY 21 ----- 20
Db 20025 ANVTYVKNQDGFVANDRPDSOGSPITIGYLIERKERNSLWVAKANDTVRSTYPCAGLV 20084
QY 21 ----- 20
Db 20085 EBLEYSFRIVALKAGSSPPSKPIEYVARKMPVDPGKPEVIDYKSVSLIMARPKHDG 20144
QY 21 ----- 20
Db 20145 GSKITGYVEACKLPDKNVRCNTAPHOIPQEEYATQLEEKAOYQPRALARTAVNISPP 20204
QY 21 ----- 20
Db 20205 SEPSDPYTLIANNPPRIDLSVAMKSLTVKAGTVYCDATVYFGKPMPTYSMKDGLLK 20264
QY 21 ----- 20
Db 20265 PAECIKMAQRMUCTLELPSVNRKDSGYTTIENS SSKSATIKLYLDRGPPASYKI 20324
QY 21 ----- 20
Db 20325 NKATSDRALSWEPLEGGSEIINYIDKRETSRPNMAQVSATVPITSCSYEKLIEGHE 20384
QY 21 ----- 20
Db 20385 YQFRICAENKYGVDVFTERPALAKNPYDPGKDPVVISNITKDHMTVSMKPRADGGS 20444
QY 21 ----- 20
Db 20445 PINGYLIERRETQAVNTEVNRKPIIERTLKAIGLOEYEFRTVAIINKAGPGKPSDAS 20504
QY 21 ----- 20
Db 20505 KAAVADPOTPPAPAPAFPRVYOTTRSSVSLSGKRAYDGGSPITIGLYEVKRADSDNMYR 20564
QY 21 ----- 20
Db 20565 CNLPQMLQKTRFEVTLMDTOYQOFRVAVNKKIGVSDPSDVPDKHYPKDILIPREGELDA 20624
QY 21 ----- 20
Db 20625 DLAKTLILBAGVTMRLYVPYKGPBPPTWSKPNVNLKDRIGIDLKSDPDTFLCENVN 20684
QY 21 ----- 20
Db 20685 KYDAGKIYIILENSCGKKEIIVYKVLDPGPPVNYTVKESKDSAYTWEPPIIDGGSP 20744
QY 21 ----- 20
Db 20745 IINYYVQKRAEKRSMSTVTTECSKTSFNVANLEBGSYFVFAENEXIGDPGETRDA 20804
QY 21 ----- 20
Db 20805 VKASQTPGPVVDLKVRSVSKSCSIGMKKPHSDGSRITIGYVVDPLTEENKQKRWKSLS 20864
QY 21 ----- 20
Db 20865 LOYSAKDLTEGKRYIFRVAENENGSGTSEIIVARDDVAPDLDLKGLPDLCTLAKEN 20924
QY 21 ----- 20
Db 20925 SNFRKIPINKAPASVSMKGEDPLATDTRVSESSAVVTLIYDQKSDAGKVTITL 20984
QY 21 ----- 20
Db 20985 KNAVIGEGTISIKVYKGPCTIPGPIKFDVTAENMTLAKNAPKDGSGSEITNYIIEKRD 21044
QY 21 ----- 20
Db 21045 SVNKKVTCASAVOKTTFRTVRLHBGMEYTFRVAENKYGVEGKSEPLVANHPRDVPD 21104

```

QY	21	-----	20
Db	21105	APPPNIVDVRHDSVLTWTDPKTGSPITGYHLEFKERSLWKRAKTPIRKURDKV	21164
QY	21	-----	20
Db	21165	TGTFEGLEFERVNAIINAGVCKPSLSEPVVALDPIDPGKPEVINITRNSVTLITWEP	21224
QY	21	-----	20
Db	21225	KYDGHKLTGYIVERKDLPSKSMKANHVNEBGAFTVTLVBSGKYEFRIRAKNTAGAI	21284
QY	21	-----	20
Db	21285	SAPSESTETICKDEYEAPTVLDPITKDLITKAGDITVLAISILGKPLPKSSWSKAG	21344
QY	21	-----	20
Db	21345	KDIRSDITQITSPTSSMLIKATRKDAGEYITATNPGTVEHYKTVLDPGPPG	21404
QY	21	-----	20
Db	21405	PVEISWSAERATLWTPLPLEDGSPKSYILKERETSRLLMTVVSBDIGSCRHATKLI	21464
QY	21	-----	20
Db	21465	QGNRIYRVSANHYKGEPVQSEPVKMDRFGPPPEKPEVSNVTKNATVSMKRPVD	21524
QY	21	-----	20
Db	21525	DGSEINGYHERREKSLRWVRAIKTPVSDLRCKVTGLDSGYEPRVSAENAGIGPP	21584
QY	21	-----	20
Db	21585	SEASDVLMDAAYPPGPPSNPTDTTKSASLAWKPHYDGLIETGYVHEHOKYGE	21644
QY	21	-----	20
Db	21645	AWIDTGTALRTQFVVPDLOTEKINERLSAINDAGVBPVPIPVEIVERMADPFE	21704
QY	21	-----	20
Db	21705	LDAELRTLVVRAGLSIRIFVPIKGRPAEYWTKNINLKNRANIENTESTILLIPEC	21764
QY	21	-----	20
Db	21765	NRVDTGKFWMTIENPAKSGEVNVRVLDTPGPVTLNRPDTIKDSVTLHMDPLIDGGS	21824
QY	21	-----	20
Db	21825	RITNVIYERKRAVRKYSTATIKCHCTYVTLSECEYFRRVMAENEYIGEPTEYIE	21884
QY	21	-----	20
Db	21885	PVKASEAPSPDSLINIDITKSTVSLAMPKPKHDGSKITGYIEAORKSDQMTHTITV	21944
QY	21	-----	25
Db	21945	KGLECVNRNLTGEBETTFQVMAVNSAGRSAPRESRPVIVEQTMPELDLRGITOKLVIA	22004
QY	26	-----	25
Db	22005	KAGDNIKVEIPVIGRPKPYTWKKGQILKQTORVNEFTATSTILININECVNSDGPYP	22064
QY	26	-----	25
Db	22065	LTAANNIVGEVDYITIOVHDIPGPTGPIKFEVSSDEVTFSWDPENDGVPTSNVYE	22124
QY	26	-----	25
Db	22125	MROTDSTTWELATVYIRTYKATRLTGLEYOFRVKAONRGVGPITSACIVANYPK	22184

QY	26	-----	25
Db	22185	VPGPCTPOVTAUVKDSMTISWHEPLSDGSPILGYHVERKERNGILMOQVSALVPGNI	22244
QY	26	-----	25
Db	22245	FKSSGLTDIAYEFRVIAENNAKSKRPSPEMLALDIPPGKPPVPLNTRHTVTLKN	22304
QY	26	-----	25
Db	22305	AKDEYTGCFKITSYIYEKRLPNGRMLKANFSNILENEFTVSGLTEDAAFEPRVIAKNNA	22364
QY	26	-----	25
Db	22365	GALSPSEPDAITCRDDVEAPKIKYDVKFQOTVILKAGBAPLEADVSGRPPIMENSK	22424
QY	26	-----	25
Db	22425	DGKELEGTAKLEIKIADFSTNLVKNKSTRDSCAVTLIATNNGEFAKHIFNVKVLDRPGP	22484
QY	26	-----	25
Db	22485	PEGPLAVTEVTSKCYLSNPPLDODGAKIDHIYOKRETSRLAMTNVASEVOVTKLKYT	22544
QY	26	-----	25
Db	22545	KLKGNEXIFRVNAVNRKYGVEPLESEPVIAVNPYGPDPKXNPBYTITTKDSNVVCWH	22604
QY	26	-----	25
Db	22605	PSDQSEIINTYIVERDKAGORWIKCNKLTLDLRYKVSGLTEGHEFERMAENAGI	22664
QY	26	-----	25
Db	22665	SAPSPTSPFYKACDIYFKRPGPNRPVLDTSRSSISIANNKPTYGGSITGYVEIALP	22724
QY	26	-----	25
Db	22725	EEDENOIVUPPAGLAKATSYITGLTENQYKIRIYAMNBSGLGEPALVGPTRKADBNLP	22784
QY	26	-----	25
Db	22785	PELELDADLRKYVTIRACCTLRLEVPKGRPAEYKWARDHESLDKASISTSSYTLI	22844
QY	26	-----	25
Db	22845	VGNVNRFDGKYLIVYENSSGSKSAFVNVVRVLDTPGPPDOLKVEYTKSVTLTMDPPL	22904
QY	26	-----	25
Db	22905	DGSKIKNYIVERKSTRKAVSTVATNCHKTSKRVDOLOEGSYFVRVLAENEGIGLPA	22964
QY	26	-----	25
Db	22965	ETAEVYKASERLPPCKITIMDYTRNSVLSWEKEHDGSHLQYIYEMQTKSGDKWAT	23024
QY	26	-----	25
Db	23025	CATVKTVEATITGLJOGEEYSPRVSANQNEKGISDPRQLSVPYLANDVIPPAPKLLFMTF	23084
QY	26	-----	25
Db	23085	TVLAGEDLKVDVPFTRPTPAVTWHDVPLKOTTRVNAESTENSLITIKDACREDVGH	23144
QY	26	-----	25
Db	23145	YVVKLNSAGAEIETLVNLVLDKPPPTGPAKMDVYTDSTILSKGPPKYDGSINNYI	23204
QY	26	-----	25
Db	23205	VEKRDSTTWQIVSATAVARTIKACRLKTCGEYOFRIAEENRGKSNVYLSNPTVAQYP	23264
QY	26	-----	25

Db 23265 FKVPDPGTPVVTLLSSDSMEVONNEPISDGSKVIGYRLERKERNILMVKLKPTIPQ 23324
 QY 26 ----- 25
 Db 23325 TKFKTGLEGEVEYEFVRSAMNIGJGKPSKSECVYARDPDPGPRPALIYTRNSVTL 23384
 QY 26 ----- 25
 Db 23385 QMKRPYDGSKITYIWEKKEPGRMKASFTNIIIDTPEVTLVEHREYFRVIA 23444
 QY 26 ----- 25
 Db 23445 AAGYSESESTGALITARDVDPRIAMDPRKYDTIVYHAGESPKVDADYKPIPTIQ 23504
 QY 26 ----- 25
 Db 23505 IKGDSELSNTARLEIKSTDEATISLVDAVRDGSNTILAKNAGERSVTVNKKVLDPR 23564
 QY 26 ----- 25
 Db 23565 GPPGPPVIGSVIAKCTLANKPPLDGGSDIINTYERRETSLVTVYDANVQILSK 23624
 QY 26 ----- 25
 Db 23625 VTKLEGNERYFRIMAVNKYGVGEPLSEPVYAKNPFVVPDAPAPVTVTKDSMIVW 23684
 QY 26 ----- 25
 Db 23685 ERPAADGSEILGYVLEKREKRGIRWTCRHLRIGELRLRVYGLIENHDEYFRVSAENNA 23744
 QY 26 ----- 25
 Db 23745 GLSESPPSAOKACDPYKKGPPNNRPVIDITRSSVFLSMKRIYDGCSEIGYIWEK 23804
 QY 26 ----- 25
 Db 23805 DVSQWMTCTPPTGINTNIEVEKLEKHEYNFRICAINAGVGEHADVPPIIVEKL 23864
 QY 26 ----- 25
 Db 23865 EAPDIDLELKIITINAGSGLFLVPIKGRPIPEYKMGVDDGIRDAIIVDTSSFTS 23924
 QY 26 ----- 25
 Db 23925 LVLDNVRNYDSGKTYLTLNENSSGFRSAFVTVRLDTPSPVNLKVTETTKDSVSIWEEBP 23984
 QY 26 ----- 25
 Db 23985 LLDGSKIKNVIWEKREAVRKSAYAVTNCNKSMLDQLQEGCSYFRTAENYIGIGL 24044
 QY 26 ----- 25
 Db 24045 PAQADPTIKVAEVPQPKIYVDVTRNSVSLSWTKPEHDSKLIQYIWEQAKHSEKW 24104
 QY 26 ----- 25
 Db 24105 SECARVKSIAQAVTNLQGEETLFRVAVNEKGRSDPRSLAVPIVAKDLVLEPDKPAES 24164
 QY 26 ----- 25
 Db 24165 SYSVQGDOLKEVPSGPRPTITWKGDLPLKOTRINVTDSLDLITLSIKETHKDDG 24224
 QY 26 ----- 25
 Db 24225 GQYGITVAVNYGOKTASIEIYTLDKPDPKGVYKPDVSAESITLSMNPPLYTGCGQJTN 24284
 QY 26 ----- 25
 Db 24285 YIVQKRDITTVWDDVVSATVARTTLKTKGTGYQERIPAEKNYGSFALSDPIVAQ 24344
 QY 26 ----- 25

Db 24345 YPKREPPGPTFAVNAISDSNVIQMHEPVNNGSPVIGYHLEKERNISLMTKVNKILI 24404
 QY 26 ----- 25
 Db 24405 HDTPQKQNLBEGIEYFRVVAENIVGKASRNSCVYARDPDPGTPPIMVKNKEI 24464
 QY 26 ----- 25
 Db 24465 TIQWTKPYDGSMTITGYIWEKRLPDGRMKASFTNIIETQFTVSGLEDDQRYEFRVIA 24524
 QY 26 ----- 25
 Db 24525 KNAAGAIKSPDSTGPTAKDEVELPRISDKPFRDTIVYHAGETFRLEADYKGRPLPTI 24584
 QY 26 ----- 25
 Db 24585 EMLRGDKETEESARCEIKTDFKALLIYKDAIRIDGGOTILASVNAKSPPVAVKIID 24644
 QY 26 ----- 25
 Db 24645 RQPPGPPVQVIGYSEKSLTQSPPLDGGSDISNYYEKRETSRLAMTVYVASEVTNS 24704
 QY 26 ----- 25
 Db 24705 LKTKLEGNERYFRIMAVNKYGVGEPLSEAPVLMKNPFVLPBPKSLEVTNIAHDSMTV 24764
 QY 26 ----- 25
 Db 24765 CQNRPDGSELIGYIWEKRDGSGIRWTKCKRRTIDRLRVYGLTDEHREYFRVSAEN 24824
 QY 26 ----- 25
 Db 24825 AAGVSPSPATVYKACDPVFRGPPPTNANIVDTKNSITLAMGRPYDGGSEILGYVE 24884
 QY 26 ----- 25
 Db 24885 ICKADEENQIYTPOTGLVTRFELISKLEHDEYKIRVCAKNGIGEATSVPYKPEP 24944
 QY 26 ----- 25
 Db 24945 KLEAPRLDSELKGIIVRAGSARIIHPKGRPTPEITWSREGEFTDKVQIEGVNY 25004
 QY 26 ----- 25
 Db 25005 TQSLDNCENDAGKYILKLENSGSKSAFVTVYKVIDTGPONLAVKEVRDSAPLWE 25064
 QY 26 ----- 25
 Db 25065 PPIIDGAKYKAVIDKRESTRKAVYANVSSKSKTSFKVENLTGAIYFRVMAENEPGV 25124
 QY 26 ----- 25
 Db 25125 GVEVETVDAKAEPPSPGKVTLPDVOSOTSASIMKEKPEHDSGLVGVEMQPKGYE 25184
 QY 26 ----- 25
 Db 25185 KMSIVAESKVCNAVVTGLSGGEYQFRVAYNEKGSDFRVIGVPIADLIIQPSLKL 25244
 QY 26 ----- 25
 Db 25245 FNTYISQAGEDLKEIIPVIGRPHRPNISWVKDGEPLKQTRVNVETATSTVLHKRND 25304
 QY 26 ----- 25
 Db 25305 DFGKTYVATNASGTAENLSVILEKPGPPVPRDEVSADEVVISMEPPATYGGCOI 25364
 QY 26 ----- 25
 Db 25365 SNIIYERKDTITTVHNVASATVARTIKITKLTGIEYQFRIPAEKNYGSAPLDSKAVI 25424
 QY 26 ----- 25
 Db 25425 VOYPERPDPGTPPTVTSISKQDMIVQMHPEVNDGGIKIIGYHLBQEKNSILMVNLKTI 25484

QY 26 ----- 25
Db 25485 P1ODTKRTTGLDEGLEEYKVSANETVIGIGKPSKVECFVANDPCDPPGRPAIVTRN 25544
QY 26 ----- 25
Db 25545 NVTLAKMKKPAYDGSKITGYIVEKKDLPGGRMKASFTVNLTEBTVVSGIVEDQRYEVRV 25604
QY 26 ----- 25
Db 25605 IARNAAGNFSEPSDSCALTARDEIDAPNASTLPKRYKDVIVHAGETVLEADIRKPI 25664
QY 26 ----- 25
Db 25665 DVTWSDKDELESTAAARMEIKSTIOKTTLVKDCIFRDGQYILKLSNWGTSKSIDPIYK 25724
QY 26 ----- 25
Db 25725 VLDRPGPEPLKVTGVTAKCYLANRPLQDGGANISMTIIRKETSRLSTQVSTEVQ 25784
QY 26 ----- 25
Db 25785 ALNKYVKRLPGNEYIFRVAVANKYIGIEPLSGPYACNPKPGRPSIPEVSATTKDS 25844
QY 26 ----- 25
Db 25845 NVVTMARPVDDGTEIEGYILERKDEGVKTCNKKTLTDLRLVYGLTGHSHYEFVVA 25904
QY 26 ----- 25
Db 25905 AENAGVGESEPSEPVFRACDALYPPGPPSNPKVVTDSHSSVSLANSKPIYDGGAPVKGY 25964
QY 26 ----- 25
Db 25965 VVEKKAADMTTCPTPLGLOGKOPTVTKIKENTENFICAINSEVGEPATLPGSVV 26024
QY 26 ----- 25
Db 26025 AQERIEPPELELDADLRKYVYLASATLRLFTYIKGRPEPEVKEKAEIILDRAGIEVT 26084
QY 26 ----- 25
Db 26085 SSETMLVLDNVTFRDGRNLTENNSSGKTAFVNVRLDSPAIVNLTIREVKDSVTL 26144
QY 26 ----- 25
Db 26145 SWEPLLDGAKITNYIVERKRETRKAVATITNNCTKTFERLENLOEGCSYFRVLASNB 26204
QY 26 ----- 25
Db 26205 YGIGLPAETTERVKVSEPLPGRGTVLDVYTNATITKNEKPSDGSKITGYVVMQTK 26264
QY 26 ----- 25
Db 26265 GSEKMSCTQVNTLEATISGLTAGEEYFRVAAVNEKGRSDPRQJGVPIANDIRIKPSV 26324
QY 26 ----- 25
Db 26325 ELPEMTFNKAROKLIDVPEGRQPAIVNKKDQTLKETRNVNSSKTVTSLIRKA 26384
QY 26 ----- 25
Db 26385 SKEDVGYELCYSNAGSITVPITIIVLDRPGPIRIDEVSCHSITISMPPEYDGC 26444
QY 26 ----- 25
Db 26445 QISNVIKEREKSTTWTWVSOAVATISIKIVLTLTGEYQFRVCAENRKYGSYSBESA 26504
QY 26 ----- 25
Db 26505 VVAEYFSPPGPIPGIKVHAHTKSTMLVTWQVFNVDGGSRTVIGYHLEKENSILMSRAN 26564

QY 26 ----- 25
Db 26565 KILLIDYQKVSGLDEGLMYEYRYAENIAGIKSKSGCEPVPARDPCDPPQPEVNTIT 26624
QY 26 ----- 25
Db 26625 RRSVLSKSRPHYDGAKITGYIVERRELJDRMKLCVNTINIOETTFVTELETDQRYEF 26684
QY 26 ----- 25
Db 26685 RYFARNADSVSEPSSESTGPIIVKDDVEPPRYMDVKFRDYIVVAKAGVLAINDIAGR 26744
QY 26 ----- 25
Db 26745 LPVISWAKDIEIERARTEISTDNHTLLVYKDCIRDTGYVLTLLKNVAGTSVAVNC 26804
QY 26 ----- 25
Db 26805 KYLDNPGPPAGPLEINGLTAEKCSLSMGRPOEDGADIDYIIVERRETSHLAWTICBGL 26864
QY 26 ----- 25
Db 26865 QMTSCVTKLLKGNETIRVTVGNKYGVGEPLSEVAITKALDPFTVPSPISLEITSVKE 26924
QY 26 ----- 25
Db 26925 SKTLCMSRPSDGSSEISGYIIRERKNSLRVAVNKAIVYDLKYSGLREGEYEYRV 26984
QY 26 ----- 25
Db 26985 YAEENAGLSLSEISPLIRADVPVLPSPSPKPIVDSKTTITIAWKPLFDGADITYG 27044
QY 26 ----- 25
Db 27045 YIVEYKSDDDIMKTSIOSLNGTEYTIISGLTGAIVYFRVSVNKGASDPSDSDPOIA 27104
QY 26 ----- 25
Db 27105 KEREPEPLDIDSEKRTLIVKAGASFTMTVPFRGAPVNTLMSKPPDLTRAYVDTTD 27164
QY 26 ----- 25
Db 27165 SRTSLITENARNDSKATTLTIONVLSAASLTIVKVLDPGPTNTVQDVITESAVLS 27224
QY 26 ----- 25
Db 27225 WDVPENDGAPVKNYHIEKRASKKAWSVTNNCNRLSYKVTNLQEGALITYFRVGENEF 27284
QY 26 ----- 25
Db 27285 GVGIPAEKRGVKITERSPPEKLGVTISKDSVSLATWIKPEHDGSRIVHYVEALENG 27344
QY 26 ----- 25
Db 27345 QKNWYKCAVASTIHVYVGLKENSEYFFRVFAENQAGLSDBRELLFVLIRKOLEPEID 27404
QY 26 ----- 25
Db 27405 MKNEPSHTVYVRAGSNLKVDPISGRPLPKVTLSDRGVPLKATMRENTETIAENLFINLK 27464
QY 26 ----- 25
Db 27465 ESVTADGRYEITTAANSGETKAFINIVLDRPGPPTGPVVISDITRESVTLKWEPPKYD 27524
QY 26 ----- 25
Db 27525 GGSQVNTIILKRETSVAVTEVSATVARTMKYMKLTTEETQFRKAKNRRGISDHD 27584
QY 26 ----- 25
Db 27585 SACTVXKLPYTPGPSPITPVNTVNTRESITVGMHEPVSNAGSAVVGHLEKDRNSITMQ 27644
QY 26 ----- 25

```

Db 27645 KANKLVRTNFKVTIISAGLIYEFVVAENAGVCKPSHPSEPVLAIDACEPRVRLT 27704
QY 26 ----- 25
Db 27705 DISKNSVLSWQOPADGSGKIYIYERDLPDGMTEASFTNVTETP11SG1LQNSQ 27764
QY 26 ----- 25
Db 27765 YEFVFAANAAGSISNPSEVGPITC1DSYGVIDLPLEYEVVYRAGTSYKLKAGIS 27824
QY 26 ----- 25
Db 27825 GRPAPTIEMWKDEKQETNALVCVENTJDLASILI1KDADLNSGCYELKLRNMGASAT 27884
QY 26 ----- 25
Db 27885 IRVQIILDKPPGPGPIEFKVTAKITLLMRPPADGAKITNYIVEKRETSRYVMSKVS 27944
QY 26 ----- 25
Db 27945 ENLEBCITTTKLIKGNELIFRVAVANKYIGSEPLSDSVANNAFVTPGPGIPEVTKI 28004
QY 26 ----- 25
Db 28005 TKNSMFWVNSRPIADGSDISGYLEKROKSLGMEKVKETLRDVRQKVTGLTENSQY 28064
QY 26 ----- 25
Db 28065 YRVCAVANAAGCPSESESEYKADPIDPPGPAKIRIADSTSSITLGNMRYVYDGS 28124
QY 26 ----- 25
Db 28125 VTGVVEIRGOEBEERTVSTKGEVRTLEYVSNLKGAVYFVSAVNAAGCEPIEMN 28184
QY 26 ----- 25
Db 28185 EPVQAKDILAPEDLDVALRTSYIAKAGDVOLIPKGRPPPTVWRDEKNLASDAR 28244
QY 26 ----- 25
Db 28245 YSIENTDSSSLITIPQVTRNDJOKYIITJENGVEPKSTVSXVLDTPAAGCKLYKHV 28304
QY 26 ----- 25
Db 28305 SRGTVTILMPPLIDGSPILINVIYERKDATKFTWSVSHKCSSTSEKILIDSEKTPFFR 28364
QY 26 ----- 25
Db 28365 RVLAEINIGICEPCETTEPVKAAEVPAPIRIDLSKSDSTSVILSRTKPPFDGGSVILEY 28424
QY 26 ----- 25
Db 28425 VYERKKGGEQWASHAGISKTCEIVSQLEQSVLEFVFAKNEKGLSDPYTIGTIVKEL 28484
QY 26 ----- 25
Db 28485 IITPEVLDSDIPGAQVTVRIGHNVHLELPYKGRKPSISWLKDGILPKESFVFRSKTEN 28544
QY 26 ----- 25
Db 28545 KITISIRAKKENGKRTVILIDNAVCHIAVPITVITGPPSKRGPIRDEIKADSVILS 28604
QY 26 ----- 25
Db 28605 WDVEDNNGGEITCYISLEKRETSQTNMKWVCSSVARTFEKPNLVKDAEQFRVABENRY 28664
QY 26 ----- 25
Db 28665 GVSQPLVSSIIIVAKHQRIRPGPKPVITVNTSDGSLJWDAPVYDGSSEVTCGHVKE 28724
QY 26 ----- 25

```

```

Db 28725 RNSILMQKVNTPISGREYRATGLVEGLDYQFRVVAENAGLSLSSPDSKFTLAVSEVDP 28784
QY 26 ----- 25
Db 28785 PGTDPYIDVTEHTITLIKRNPLRDGSKLYGYSLEKROGNERNVRCNFTDVSECQTVWG 28844
QY 26 ----- 25
Db 28845 LSPGDRYEFRLIARNAVGTISPQSGCIWTRDENVPPIVEFGPEYFDGLIKSGESLR 28904
QY 26 ----- 25
Db 28905 IKALVQGPVPRVTFWKDGEIEKRNNEITDVLGTSLEFVADTRDNKGYVYEAKNAS 28964
QY 26 ----- 25
Db 28965 GSARAEIKVQVDTPGRVYGP1RFTNITGEKTLMDAPLNDGAPITNYIIEKRETSRL 29024
QY 26 ----- 25
Db 29025 AWALIEDKCEASVTAIRL1NGNEVQFRVSAVNRFGVGRPLDSPVVAQIQYTVDPAGI 29084
QY 26 ----- 25
Db 29085 PEPSNITGNSITLTVARPESDGSEIQYIIEKREKSTRVYKVISKRPISETRFKVTGL 29144
QY 26 ----- 25
Db 29145 TEGNEYEFHMAENAGVAPASGISRLIKCHEPVNPPGPTVAVKVTDSKTTVLEMSKP 29204
QY 26 ----- 25
Db 29205 VPDGMEIIGYIIECKADLDPMKNVNAECYKRTVTVDLQAGEYRFRVSAINGAKG 29264
QY 26 ----- 25
Db 29265 DSCVYTG1KAVDRILTAPEDIDANFKQTHYRAGASIRLE1AVQGRPTPAVMSKPSDN 29324
QY 26 ----- 25
Db 29325 LSLRADINTDSFTLVENCNNDAGRTTLTVENNSGKSITFTVKVLDPPGPGPITF 29384
QY 26 ----- 25
Db 29385 KDVTGSGATLMDAPLDLGGARINHYVEKREASRHSQVISEKTRQIEFYNDLABGP 29444
QY 26 ----- 25
Db 29445 YFRVSAVNEVGVPEYEMPEPIVATBQPPRRLDVTDSKSAVLAMLPDHDGSR1 29504
QY 26 ----- 25
Db 29505 TGYLLEMQKGSDFVEAGNTKQLTFTVERLVEKTEYFVFAKANDAGYSEPREAFSSVI 29564
QY 26 ----- 25
Db 29565 IKEPQIEPTADLTGITTNOLITCKAGSPFTIDVPISGRAPKVTWLEEMRLKETDRVIT 29624
QY 26 ----- 25
Db 29625 TYNDRTTLTVKDSMGDSGRFLENTAGVTFVTVVVYIGRPGVPIEVSVAES 29684
QY 26 ----- 25
Db 29685 CVLSWGEPRDGGTETITNYVEKRESGTTAMOLVNSVAKFOIKVTHLTKTMEYSEVSS 29744
QY 26 ----- 25
Db 29745 ENRFGYSKPLBESAPITAEHPVPPSAPTRPEVYHVSANAMSIRMEEPYHDGSGKIIGTW 29804
QY 26 ----- 25
Db 29805 EKERNTILVYKENVPCLECNVYKTVGLVEGLEVOFRVYALNAGVSRASASRPIMAQN 29864

```


QY 26 ----- 25
Db 29865 PVDAPGPEYTDVTRSTVSLIWSAPAYDGSKVYGLIERKPVSEVGDHMLKONTIYS 29924
QY 26 ----- 25
Db 29925 DNEFTYVALSRGDYEFRRVLAKNAGVISKSGSESTGFTCRDEYAPPKAELDARLHGDLY 29984
QY 26 ----- 25
Db 29985 TIRAGSDVLDAAVGAPPEKLIWTKGDELCEKVSLOYTGKRATAVIKFCDRSDSGK 30044
QY 26 ----- 25
Db 30045 YTLTVKNASGTAVSVNWKVLDSPGCKLTVSNVTOEKTSLMSLPQEDGAEITHYIV 30104
QY 26 ----- 25
Db 30105 ERRETSRLNNVIVEGCEPILSYVVTIRLKNNEYIFRYRANWKYGPVVESEPIVANRNF 30164
QY 26 ----- 25
Db 30165 TIBSPGCIPEEVGKEHEIIQWTKPESDGNELSNVLVDKREKSLRMTRVNKDYVVD 30224
QY 26 ----- 25
Db 30225 TRUKVTSLMEGCDYQFRVTAVNAGNSEPSEASNFIISCREPSYTPGPPAPRVVDYTKHS 30284
QY 26 ----- 25
Db 30285 ISLAWTRPMTDGTDIYGVYLEMEQKOTDQWTRVHNATIRNTEFTVPDLKMGOKTSFRY 30344
QY 26 ----- 25
Db 30345 AAVNVKGMSEYSESLAIEFEVERIEIPDLIELADLKTITIRAGASILRLWVSYGRPPY 30404
QY 26 ----- 25
Db 30405 ITWSKQIDLASRAIJDITTESILLYDKVNRVDAQVYTIABNGSKKSAVLVRYDT 30464
QY 26 ----- 25
Db 30465 PGCPGVKVEVSRDSTVITWEIPTIDGAPVANNIYERKRAAMRAFKTYTKCSKTLR 30524
QY 26 ----- 25
Db 30525 ISGLVEGTMKXFRVLPENTYIGPEPETSDAVLVSEVPLVPANLEVVDTKSTVTLAMEK 30584
QY 26 ----- 25
Db 30585 PLYDGSRLGVYLEACKACTERKMKVYLKPTVLEHTVTSLNEGHOYLFRIBAONEKV 30644
QY 26 ----- 25
Db 30645 SEPRETVAVTQDLRVLPTIDLSINPKTIHVPAGRPELVPIAGRPAPASWFEAGS 30704
QY 26 ----- 25
Db 30705 KLBESERVVEITHKVAKLPIRETTIRDTGEYTLLEKNVTGTTSETIKVILIIDKBPPTG 30764
QY 26 ----- 25
Db 30765 PIKIDEIDATSIISWPELBDGAPLGSVVEQNDARHPGMLPVSESVTRSTFKFTRLT 30824
QY 26 ----- 25
Db 30825 EGNERYFRVAANRFGISYLOSEVTECRSSIRIPGPELTIQIFVSRDGMTLTWPPED 30884
QY 26 ----- 25
Db 30885 DGSQVGTGIVIERKEVRADNRVAVKVPVYTRYSSTGLTBELEYEHRVTAIINAGSGKP 30944

QY 26 ----- 25
Db 30945 SRESPEIVANDPIAPPGKQPNRVYDTFRTSVLSMVSPEDEGSKRVYGLIEMQKVDH 31004
QY 26 ----- 25
Db 31005 EWTCKNTPTKIRETYTLHLPOGAERYFRVLACNAGBGEPAVPGVATEMLEYPPYE 31064
QY 26 ----- 25
Db 31065 LDERVOEGIEVRQGVYIRLPIPIKPPICKWTEGGDISKRAMIASEHTLVIKEA 31124
QY 26 ----- 25
Db 31125 DRGDSGYDVLLENKGRKAVYIKVRVIGSPNSPGPIEYDIOVRSYRVSNRPADGG 31184
QY 26 ----- 25
Db 31185 ADILGYILERRRVPKAMWTIDSRVRGSTLVYKGLKENVEYHFRVSAENQGISKPKSE 31244
QY 26 ----- 25
Db 31245 ERYTPKTIAPPEPSNPPEVLDVTKSSVLSWSXPKDDGSRVYGYIERKSTDKWV 31304
QY 26 ----- 25
Db 31305 RHNKQITTTMYTGLVPEAEYQFRIIAQNDVGLSEISPASEPVYCKDPDKPSQBEL 31364
QY 26 ----- 25
Db 31365 ELLISKDSVTLQWKEPBCDGKEILGYWEYRQSDSAMKSNKRIKDKOPTIGLLE 31424
QY 26 ----- 25
Db 31425 AVEYERVAENETGLSNRRTAHSIKTKLTSGEAPGIRKEMKDVYTLGPAQLSCQIV 31484
QY 26 ----- 25
Db 31485 GRPLDILKWRFGKELIQSRKYMSSDGRTHLIVWFEQDEGEVYCIATNEVEVETS 31544
QY 26 ----- 25
Db 31545 SKULLQATPOHPGYPLEKYYGAVSTLRHWMYIGHPAMTFHQLLONSENTI 31604
QY 26 ----- 25
Db 31605 ENTEHYTLVKNQKQKTHAGKYKQVLSNPTGYDAILDVEIIOKPKPTGPIYIEMALK 31664
QY 26 ----- 25
Db 31665 NSAVISMNRPADGGSMNTNVVENKCEAKGAEQVSSAISVYTCRVLNENAGYFR 31724
QY 26 ----- 25
Db 31725 VSAONTFGISDPLEVSSVITIKSPREKAPGKPTITAVTKDSQVAMKPPASDGAKIR 31784
QY 26 ----- 25
Db 31785 NYLEKREKQKQKWSVYTHEIRFVSVKMLIDGLEEFVKNCENLGSEHSEISEPI 31844
QY 26 ----- 25
Db 31845 TPKSDVPIQAPHFKEELNLRVYOSNATVCKYTGHPKPIVKYRQGEIILADGLKRYI 31904
QY 26 ----- 25
Db 31905 QEFKGYHQILLASVYDDATVYQVRATNCGSVSGTASLEVEVPAKIHLPKTLEGAV 31964
QY 26 ----- 25
Db 31965 HALNGEVYSIKIPSGKRPVITWQKQDILDNNGHYQVATRSFTSLVFPNGVERDAG 32024
QY 26 ----- 25

```

Db 32025 FYVYCAKNREGIDOKTEVLDVADVPDPBGVYKSDVSBDVNLJTTEPASDGSKITNTYI 32084
OY 26 ----- 25
Db 32085 VECATTAERMLNVOGARETRYVNLFGKTSYOFVIAENKRGSKPSEPSEPTIKED 32144
OY 26 ----- 25
Db 32145 KTRAMNDDEVEDTREVSMTKASHSSTKELYEKYMAEDLGRCGCIYHRCVETSCKTY 32204
OY 26 ----- 25
Db 32205 MAKVVKGTQDYLAKKEISILNIAHRNLIHESFESMEELWIFEFISGLDIPERIN 32264
OY 26 ----- 25
Db 32265 TSAFELNEREIVSYHVOCBALQFLSHNIGHFDIREPNIYQTRRSSTKIIEFGQARQ 32324
OY 26 ----- 25
Db 32325 LKQDNFRLIFTAPETAYAPRHQHDVSTATDMWSLGTLYVYLLSGINPLAETNOIIE 32384
OY 26 ----- 25
Db 32385 NIMNAEYTFDEEAFKEISIEAMDVDRLLVKEKRSMTASEALQHPMLKQKIERVSTKYI 32444
OY 26 ----- 25
Db 32445 RTLKRRRYHTLLIKKIDINMYSAARISCGAISQKGVSAKVASIEIGVSGQIMHA 32504
OY 26 ----- 25
Db 32505 VGEKGVHVKYCKLENDQSTQYTWYGVQLENSEKKEITTEDGVALLVYKQITKLDG 32564
OY 26 ----- 25
Db 32565 TYRCKVNDYGEDSYAELEFVKGVREYDYCRRTMKIKRRDTRMLERPEFTLPLY 32624
OY 26 ----- 25
Db 32625 NKTAIVGENVRFGVTTVHPPEHVTWYKSGQKIKPGDNKKYTESDKGLYOLINSVTT 32684
OY 26 ----- 25
Db 32685 DDAEYTVVANKYGEDSKAKLTVTLHPPTDSTLRPMFKRLLANAQOGQSVCEIR 32744
OY 26 ----- 25
Db 32745 VSGIPPTLWKWDGQPLSGPNIIEIHGIDYALHIRDTLPEDNGYRVATNTAGST 32804
OY 26 ----- 25
Db 32805 SCQHLQVERLRYKQEFKSEHERHVOKIDKTLMAELISGTSEVPYLQVAKELRE 32864
OY 26 ----- 25
Db 32865 AAVLYKPAVSTKVKGFEKLEIEKKERKLRMPDYVPEPRKYQOTTIEDQRIKQFVPM 32924
OY 26 ----- 25
Db 32925 SDMKWYKIKQDYEMPGELDRVQKRPKRLKLSKWEQTYVAPLPRITDQYRPAKWRIPKLS 32984
OY 26 ----- 25
Db 32985 QDDLEIYRPAKRTPSPODYFYPRRRLSGDISDEELLPLIDIDYLANKKEEERLALBE 33044
OY 26 ----- 25
Db 33045 ELIEIGSASPPSRBPHELSLRYSSPOAHVKEETRKDRYSTYHILPTKAEASTYAE 33104
OY 26 ----- 25

```

```

Db 33105 LREHQAAYRQPKORIKAREDEDELLRPVTTQHLSEYKSELDMSEKRSRKR 33164
OY 26 ----- 28
Db 33165 QREVTETIEEYHISKHAORESSSSSKRLRRRSISPTIILMPVSELIRSPQPA 33224
OY 29 ----- 28
Db 33225 EEEEDTERSPTEPRTNRPSPVSSERSLSREPSARFDIFSRYESMKALKTQTS 33284
OY 29 ----- 28
Db 33285 RRYEVLSCQPELTDHAPRTLRMSHRVPCQONTFTLINOQKPTAEKTYHNGVLOES 33344
OY 29 ----- 28
Db 33345 SKIHTNTSGVLTLEILDCHDDSGTYRACVTNKGASDYNATIDYTGQDITYASQRD 33404
OY 29 ----- 28
Db 33405 EEVPSVPPELTREAYAVSSFKTESEMASSVYEVKSQMTETRESISYHSASAEK 33464
OY 29 ----- 28
Db 33465 SALEEKSLERSTTRKIKITLAAIRILTPRSMYVEGESARFSCODIDGEPVYTMK 33524
OY 29 ----- 28
Db 33525 GOYLSTARHQVTTKYSSTPEISVQASDBGNYSVVENSEGQOAEFTLIQKARTE 33584
OY 29 ----- 28
Db 33585 KAVTSPPRVKSPPEPVKSPKAVKSPKRVKSPESHKAVSPTEKPPTEKVOHLPVSAP 33644
OY 29 ----- 28
Db 33645 PKITQFLAKASKELAKLTCVYESVYLAKEVYTKDKKILAKENHQPYHYSADQTYELK 33704
OY 29 ----- 28
Db 33705 INNLESQGEYVCHSGTSGTNTLQFMQOAFKSIHEKYSISETKSDQKTESTVT 33764
OY 29 ----- 28
Db 33765 RTEPRAPRIPSSKPIVYTGLODTVSSDSYAKFAVKATGEPRTAIWTKDGKAITQGGK 33824
OY 29 ----- 28
Db 33825 YKLSKQGFLEIHKTDSDGLYTCYKNSGVSSSCKLTIKAITDEAKVSTQKT 33884
OY 29 ----- 28
Db 33885 SEITPOKKAVOEISOKALRSHEIKMSKQEKILAKESKVLISEYKSAATILE 33944
OY 29 ----- 28
Db 33945 KSTIREIITQTSQASEEVTHAIEIKAFSTQMSINEGQRLVLANTAGATDYKWLINGVEL 34004
OY 29 ----- 28
Db 34005 TNSERYRGVSGSDQTLITKQASHDEGLTJCISKTBGIVACQYDPLTSLKELSDAPAFI 34064
OY 29 ----- 28
Db 34065 SQPRQINQGNVLTCEISGEPSELEMFKNLPLISISHVSISSRNVYSLERINAS 34124
OY 29 ----- 28
Db 34125 VSDSGYTTIKANFNKQCSATASLAWLPLVEEPSREVLKTSQDTSLOGSFSSQSVQNSA 34184
OY 29 ----- 28
Db 34185 SKQEASFSSPSSSASMTKMFASMSQMSMSQESFVEMSSSPFKGISNTQLESSTS 34244

```

OY 29 -----GRF 31
 Db 34245 KMLRNGINGIPKIEALPSIDISIDEGVLTVCACAFGEPEPEVWVSCGGRKHISBOGRF 34304
 |||

RESULT 5

0924X6 PRELIMINARY; PRT; 7463 AA.
 ID 0924X6
 AC 0924X6
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE CNA peptide synthetase I.
 GN CNAPE1 OR SCO3230 OR SCO63.03C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
 OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE=97000351; PubMed=8843436;
 RX Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Barris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittowles E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrett B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 DR EMBL AL035640; CAB38518.1; -.
 DR HSSP: P14687; IAMD.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; Condensatn.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003880; Peptide_attach.
 DR InterPro: IPR000408; Reg_chf_condens.
 DR InterPro: IPR000169; SHProl_acsite.
 DR Pfam: PF00501; AMP-binding; 6.
 DR Pfam: PF00668; Condensation; 8.
 DR Pfam: PF00550; pp-binding; 6.
 DR PROSITE: PS50075; ACP_DOMAIN; 6.
 DR PROSITE: PS00455; AMP_BINDING; 6.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
 DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
 KW Phosphopantetheine.
 SO SEQUENCE 7463 AA; 798598 MW; 6A168F63D4CFED54 CRC64;

Query Match 42.9%; Score 67; DB 16; Length 7463;
 Best local similarity 0.5%; Pred. No. 3.3e+03;
 Matches 17; Conservative 2; Mismatches 8; Indels 3127; Gaps 3;

OY 4 HXSMEXRTP----- 13
 Db 3537 HWSVLDGTPPTLRDLPARDTHATAGEITAELEDDTTEALLTWVPGVCHAVNDVFLSTF 3596
 OY 14 ----- 13
 Db 3597 ALAVAGRRGRGDADAPVYLLDSEHGHEAVPGVELSRTAGWFTSMYVRIAPAGAS 3656
 OY 14 ----- 13
 Db 3657 GDGSALRALKAVEQLRTVPGDGLGYGLLRHINPRTALALPLPERGFNYLGRIGQEG 3716
 OY 14 -----DIN----- 16
 Db 3717 TDEAPWTIEGGVAGIDGAPPLAHPVDVNAVARETADTGLRARWYSRTALEPEDTQL 3776
 OY 17 ----- 16
 Db 3777 ADTWRLRLRYBEARQPGAGGILTPSDIAHPALADEIEDLEHTVPGLDIPLAPLOEG 3836
 OY 17 ----- 16
 Db 3837 FLPLNYDENARDVYVQGLAPFLDESPDGRMAAGALLRRHANLRAGFROTATGTWYQ 3896
 OY 17 ----- 16
 Db 3897 VVPALPEPMRECDLITDRADEERDAENGRLAAGDERERFDTLSPLNRTAIRLSADRV 3956
 OY 17 ----- 16
 Db 3957 RLWMTNHHLLDGRSMPLLMQELTBYLSSGGDPVSLPPRYRDIHLNLAGARDANDADA 4016
 OY 17 ----- 16
 Db 4017 NRRSLGIDEATLILAPDAGPAEAPLGIHFGIDRDATALSMARGVNTWVVGAMA 4076
 OY 17 ----- 16
 Db 4077 LALAQATGDDVFGATVSGRPPELPGVESMIGLFINLPVARILDOAEPLGLDERRLQN 4136
 OY 17 ----- 16
 Db 4137 PQARLLDHQPLADIQHWAGGELPDTAMVQNPVPSADTTSRODLRYAGTDAVEST 4196
 OY 17 ----- 16
 Db 4197 DPAVNLVAFTRDMLRLRLDYRADACADLYRSLADRMVYLEALVTDSDRPVHLDTLD 4256
 OY 17 ----- 16
 Db 4257 PAVRERVLVENGAPQLPGTPLHELISBOARLIPDAAVYVCDGTSLYAEIDGAGOLA 4316
 OY 17 ----- 16
 Db 4317 RHLLSEGIGAEEDFVALAKSLIDAVISMLAVLKTGAATPLIDPPYPAERITVMDLARA 4376
 OY 17 ----- 16
 Db 4377 LTLTEPVERVYTGSVATVDEERRSPWSARHAAVMTYSGTGRPGVYIEHMLATY 4436
 OY 17 ----- 16
 Db 4437 LHRANNTYAMGVYVHSPPLAFDITLALNTPLTSGTGHLSIESDTPSLIKATPS 4496
 OY 17 ----- 16
 Db 4497 RLPLLTTLPEASPSHTLLIGLEALHTDHLVWRTQHPGVQIINAYGTESTVNTDHNH 4556
 OY 17 ----- 16

```

Db 4557 GEDRPGDVPVIGRPFANTQYVILDSALRPVAPGVGTGEIYIAGBQLARGYLGRPALTAERF 4616
QY 17 ----- 16
Db 4617 TANPHSTPGAMRYRTGDLAHMNHGHLYDGNADHOIKIRCHRIEPELEATLTQAOTCI 4676
QY 17 ----- 16
Db 4677 TQATVQAREDPGDORLWALVYVNDSTEYDEKTYROALTSALPDYWPVSALVTLDALEPL 4736
QY 17 ----- 16
Db 4737 PNGKIDRTALPAPASASTAGRAPRTRPREBVLCTIFAEVLGVDLVTIDNFPDIGHSL 4796
QY 17 ----- 16
Db 4797 AWRLVSTRALGVELSTIROLFETPTVAGLAELDASGVETRTALTARPERILPLSTAOQ 4856
QY 17 ----- 16
Db 4857 RLMELHJEGBSATYNTVTLRLGGLVDALNAISDVVAHRESLRTVTEDEBGAYOI 4916
QY 17 ----- 16
Db 4917 VLPVEASTPTVVVAEEIIGDRLDEAVGHCEDLAQELPAKUSLEKVSEREHVLLELH 4976
QY 17 ----- 16
Db 4977 HIASDANSRAPLAODLTAAATAAARSEAPKAPLTVQYATAMQOETIGDDTDADSLAG 5036
QY 17 ----- 16
Db 5037 KQLAWMKQALGLPEQIDLPDTHPRPAVAGYSGDRVPFTVTEHLRTLELATWTSAF 5096
QY 17 ----- 16
Db 5097 MVIOAAVAVILLTRLAGEDIPIGTPVAGRTDADADLIGLFINTLVLRITDTSQDPFERRL 5156
QY 17 ----- 16
Db 5157 LDRVRDIDLAAYARQDLPFERLVEALNPARTLSHHPLFQYULFENNDHGALKDISELP 5216
QY 17 ----- 16
Db 5217 GLTVALKREVQRTSSKFDLSFGFAESFDSRRQGLEALDSTELDRSAQAIADRELR 5276
QY 17 ----- 16
Db 5277 VLEAVTAPDRPIGAVELMDPAERERVLVEKNGAPQOLPGTPELHISEQARLTPDAVAV 5336
QY 17 ----- 16
Db 5337 VCDGTTLYAELDRRANQLANHLIGBGLGAEDEVALALAKSLDAVISMLAVLKGAAYLP 5396
QY 17 ----- 16
Db 5397 IDPYPAERITVMDAQAPLTLTAP1PPASYDSRPTSEITDYERSPWSARHAYMYT 5456
QY 17 ----- 16
Db 5457 SCSIGRPGVYTEHHALATYIARARNTYTAATGTVLHSPILAPLTYTALMPTLACGYV 5516
QY 17 ----- 16
Db 5517 HLTSLERAEVQPSLIKAPSHLPILTLTPETASPSHTLILGGFALHTDHLATWRTQPGA 5576
QY 17 ----- 16
Db 5577 QIINAYGPTESTVNTDHHVSEDTPDGPVPIGRPFANTQYVILDSALRPVAPGVGTGEIYL 5636
QY 17 ----- 16

```

```

Db 5637 AGEQLARGYLGRPALTAERFTANPHSTPGARNTRTGDLAHMNHGHLYTDGNADHOIKL 5696
QY 17 ----- 16
Db 5697 RGHRIEPELEIETTLQAOTGTQVQLREDTPGQRLVAVLVVNDSTEYDEPTLRDALAS 5756
QY 17 ----- 16
Db 5757 ALPDYWRPSAVVTLIDALPULTPNGKIDRTALPAPASASTAGTGRTPREELCTIFAEVL 5816
QY 17 ----- 16
Db 5817 GVDLVTIDNFPDIGHSLAFLVSRARIALGVELSKOFEPEPTTIGLSGARDRAGRA 5876
QY 17 ----- 16
Db 5877 RAALTAARPERIRIPSTANOORLFLHJEGBSATYNTVTLRLGGLVDALNAISDVVAHRESL 5936
QY 17 ----- 16
Db 5937 ARHESLRTVTEDEBGAPROVIHAWEPCHLPLGVVDTGGEIDAMLGAGVHAFDLTAGIP 5996
QY 17 ----- 19
Db 5997 VRAVLEFRISQREHVLLELHITATDANSRTPLGHDLLAAYSACAGADVAPANEPLPVQYAD 6056
QY 20 ----- 19
Db 6057 YALMQREVLDGSDADAPAGRLAYWTRQIDLPLBQDLPDTRPAPAVASQGDVRVAFSL 6116
QY 20 ----- 19
Db 6117 DADLYVRLTELARAHSTTFVWQALAVLLTRLGAGEDIPIGTPVAGRTDADENLVGF 6176
QY 20 ----- 19
Db 6177 FYWTVLVRNDTSGNPTFERLLETRTDLAAYAHQDLPFERLVEALNPARTLSHHPLFQY 6236
QY 20 ----- 19
Db 6237 MLILSTAETPDASLALPLGLRVAERSRLGAKVDALEAALAEVRDGEKSTIGLGLDPR 6296
QY 20 ----- 19
Db 6297 TDLFDRSTARSLVERFRTLEAVVADPGVRLSRFVLTGSEKRSILDRGTGPLEGLDAT 6356
QY 20 ----- 19
Db 6357 LPBELFAEQALRTPGAPALVKGSTVSYAELDLRTNRLARLLRQGVRRGTPVVMLEKSP 6416
QY 20 ----- 19
Db 6417 AHVATLALAKGAVVPLEDTYPLDRMRHVAVDTAAVLITLDRAEARAGOLGARVAV 6476
QY 20 ----- 19
Db 6477 DERGAAPSGSEADAPGTGTGTGTSRSGYVUDAPAEVGLRFDOLAYVWTSSTGVPRGV 6536
QY 20 ----- 19
Db 6537 AVTHRGVNDLVRDHCWRPGVHERVLLHAPHADVSCYEMKVVPLVSGVWVAPPGHLDPA 6596
QY 20 ----- 19
Db 6597 AITDLTAADITAIHLTLAGFRRVAEAECPAGVREVLIGDVPVSPAARVAVLHHPRI 6656
QY 20 ----- 19
Db 6657 VLRHLXGPTETTLCTQHEVTAPYEAKGSLPVGH 6690
QY 20 ----- 19

```

RESULT 6
000308

```

ID 000308      PRELIMINARY;      PRI;      870 AA.
AC 000308;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE WMP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97313427; PubMed=9169421;
RA Pirozzi G., McConnell S.J., Uveges A.J., Carter J.N., Sparks A.B.,
RA Kay B.K., Fowkes D.M.;
RT "Identification of novel human WW domain-containing proteins by
RT cloning of ligand targets."
RL J. Biol. Chem. 273:14611-14616(1997).
DR EMBL: U96114; AAC51325.1;
DR HSSP: Q1326; 1PIN.
DR InterPro: IPR000008; C2.
DR InterPro: IPR000569; HECT_domain.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR Pfam: PF00632; HECT; 1.
DR Pfam: PF00397; WW; 4.
DR PRINTS: PR00403; WMDOMAIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM00119; RECTG; 1.
DR SMART: SM00456; WW; 4.
DR PROSITE: PS50237; HECT; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 4.
DR PROSITE: PS50020; WW_DOMAIN_2; 2.
SQ SEQUENCE 870 AA; 99070 MW; 63342143211D0294 CRC64;

Query Match
Best Local Similarity 42.3%; Score 66; DB 4; Length 870;
Matches 18; Conservative 2; Mismatches 11; Indels 588; Gaps 4;

QY 1 SRXHXHS----- 7
DB 191 SRHRHSGASARTPATGSGSPGARSRRHROPVKNSHGSLANGTYNDEPTATDPEPSV 250
QY 8 ----- 7
DB 251 VGVTPPAAPLSTVPNPTSLPAPAPAECEPSTSGIQQLPAAQAPDALPAGWQRE 310
QY 8 ----- 12
DB 311 LPMGRVYVDHNTKTTTWERPLPPGMEKRTDPRGRFYVDHNTRTTQWRPTAEVRYNTE 370
QY 13 ----- 12
DB 371 QMSQRNOLGAMQRPQRLYQFNSASTDHDPLGLPPGMEKRODNGRYVYVNHNTTT 430
QY 13 ----- 26
DB 431 QWEDPTQGMIGEPALPPGMEKRTSEGVRYFVDHNTRTTTFKDPKPGFESGTGQSGPCA 490
QY 27 ----- 26
DB 491 YDRSFRMKYHOFPLCHSNALPSHVKISVSRQTLFEDSFQIIMMKPYDLRRRLYYIMRG 550
QY 27 ----- 26
DB 551 EBGIDYGIAREWFFLLSHVILNPMYCLFETAKNNYCLQINPASSINPDHLYTFRPIGR 610
DB 611 FIAMALHGFIDWGPFLPYKRLNKRPTLKDLSEIDPEFYNSIWKENNIEGGLTEL 670
QY 27 ----- 26

```

```

DB 671 YFIQMDLIGKTYTHLKEGESIRVTEENKEEYIMLTDKRFTRGVGEQTKAFIDGFNE 730
QY 27 ----- 26
DB 731 VAPLEWLFDEKELELMCGMEIDMDSQSTFYRHYIKNSKQIQMFQGVAKMENEK 790
QY 27 ----- 31
DB 791 RIRLLQFVYTGCRLEVGSE 809

RESULT 7
ID 096CZ2      PRELIMINARY;      PRI;      870 AA.
AC 096CZ2;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Medd-4-like ubiquitin-protein ligase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC013645; AAH13645.1;
DR InterPro: IPR000569; HECT_domain.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR Pfam: PF00632; HECT; 1.
DR Pfam: PF00397; WW; 4.
DR PRINTS: PR00403; WMDOMAIN.
DR PROSITE: PS50237; HECT; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; UNKNOWN_4.
DR PROSITE: PS50020; WW_DOMAIN_2; 2.
DR KW
SQ SEQUENCE 870 AA; 98912 MW; FCCD75CBA61F2204 CRC64;

Query Match
Best Local Similarity 42.3%; Score 66; DB 4; Length 870;
Matches 18; Conservative 2; Mismatches 11; Indels 588; Gaps 4;

QY 1 SRXHXHS----- 7
DB 191 SRHRHSGASARTPATGSGSPGARSRRHROPVKNSHGSLANGTYNDEPTATDPEPSV 250
QY 8 ----- 7
DB 251 VGVTPPAAPLSTVPNPTSLPAPAPAECEPSTSGIQQLPAAQAPDALPAGWQRE 310
QY 8 ----- 12
DB 311 LPMGRVYVDHNTKTTTWERPLPPGMEKRTDPRGRFYVDHNTRTTQWRPTAEVRYNTE 370
QY 13 ----- 12
DB 371 QMSQRNOLGAMQRPQRLYQFNSASTDHDPLGLPPGMEKRODNGRYVYVNHNTTT 430
QY 13 ----- 26
DB 431 QWEDPTQGMIGEPALPPGMEKRTSEGVRYFVDHNTRTTTFKDPKPGFESGTGQSGPCA 490
QY 27 ----- 26
DB 491 YDRSFRMKYHOFPLCHSNALPSHVKISVSRQTLFEDSFQIIMMKPYDLRRRLYYIMRG 550
QY 27 ----- 26
DB 551 EBGIDYGIAREWFFLLSHVILNPMYCLFETAKNNYCLQINPASSINPDHLYTFRPIGR 610

```

```

OY 27 ----- 26
DB 611 FIAMALYHNGKFIPTGFTLIPFYKRMNKRPTLKDLSIDPEFYNSIWIKENNLEBCEGL 670
OY 27 ----- 26
DB 671 FTIQDEILGKVTTHLEKGEISINVTENKEEYIMLLTDMRTGVEBQTKAFLDGNE 730
OY 27 ----- 26
DB 731 VAPLEMLRTFDEKELEMLGCMQOEIDMSDMOKSTIYRYTNSKOIQOFNQYKEMDNK 790
OY 27 -----PGRF 31
DB 791 RIRLQFVGTCTCLPVGPF 809

RESULT 8
OQDBH0
AC OQDBH0; PRELIMINARY; PRT; 870 AA.
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 1300010006RIK protein.
GN 1300010006RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi D., Fukuda S.,
RA Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., NIKaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Steudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baidarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Bayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004962; BAB23702.1; -
DR BSSP; Q13526; 1P1N.
DR MGD; MGI:191414; 1300010006RIK.
DR InterPro; IPR000008; C2.
DR InterPro; IPR000569; HECT_domain.
DR InterPro; IPR002349; WW.
DR InterPro; IPR001202; WW_Reps_WMP.
DR Pfam; PF00632; BECT; 1.
DR Pfam; PF00397; WW; 4.
DR PRINTS; PR00403; WMD0ANIN.
DR SMART; SM00239; C2; 1.
DR SMART; SM00119; HECT; 1.
DR SMART; SM00456; WW; 4.
DR PROSITE; PS50237; HECT; 1.
DR PROSITE; PS50119; WW_DOMAIN; 1; 3.
DR PROSITE; PS50020; WW_DOMAIN; 2; 2.
DR PROSITE; PS50020; WW_DOMAIN; 2; 2.
SQ SEQUENCE 870 AA; 98760 MW; 72B34A1B727A7FB3 CRC64;

```

```

Best Local Similarity 2.9%; Pred. No. 1.3e+02;
Matches 18; Conservative 2; Mismatches 11; Indels 588; Gaps 4;

OY 1 SRXKHS----- 7
DB 191 SRTHRHSGSARATATASEQSPARNRHOPVANSSSGLANCTVNEEPTPAPEPSSSV 250
OY 8 ----- 7
DB 251 VGVTSLEPAALSVSNPNTSLPQSTPAEGEASTSGTQQLPAAQADALPAGNEORE 310
OY 8 -----MEKRT----- 12
DB 311 LPNGRVYYVDHNTKTTWERPLPPEWEKFRDPGRFRYYVDHNTRTTTPAETAYANE 370
OY 13 ----- 12
DB 371 QMSORNOLOGAMOBFSQRFLOYSSASSTDHDLGLPLPPEWEKRODNGRVYVNHNTRTT 430
OY 13 -----PDINPAM---YXXKGR----- 26
DB 431 QWEDPRQGMIOBPALPPEWEKMYTSEGYYVFDHNTRTTTPKDPREGSGTKQSPGA 490
OY 27 ----- 26
DB 491 YDRSFRWKYHQFPLCHSNALPSHVKLSVRSQTLFEDSFQIIMMKPYDLRRRLYIMNG 550
OY 27 ----- 26
DB 551 EEDLDGGIARENFFLSBEVLNPTCLPEYAGKNKYCQIINPASSINDHLYTFRTIGR 610
OY 27 ----- 26
DB 611 FIAMALYHNGKFIPTGFTLIPFYKRMNKRPTLKDLSIDPEFYNSIWIKENNLEBCEGL 670
OY 27 ----- 26
DB 671 FTIQDEILGKVTTHLEKGEISINVTENKEEYIMLLTDMRTGVEBQTKAFLDGNE 730
OY 27 ----- 26
DB 731 VAPLEMLRTFDEKELEMLGCMQOEIDMSDMOKNATRYHTKSKOIQOWEYKEMDNK 790
OY 27 -----PGRF 31
DB 791 RIRLQFVGTCTCLPVGPF 809

RESULT 9
OQ4H07
AC OQ4H07; PRELIMINARY; PRT; 1889 AA.
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Putative phosphatidylinositol kinase.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=4530;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Mofiat K.S., Hill J.N., Gansberger K.,
RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tslirin T.,
RA Riggs F., Hsiao J., Zisman V., Blunt S., Pal G., Vanaken S.E.,
RA Uterback T.R., Feldlyum T.V., Quackenbush J., Salzberg S.L.,
RA White O., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNBa001.8H01 genomic sequence.";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC087181; AAK30486.1; -
DR InterPro; IPR001263; P13Ka.

```

Query Match 42.3%; Score 66; DB 11; Length 870;

DR InterPro: IPR000403; P13_P14_kinase.
 DR Pfam: PF00613; P13K; 1.
 DR Pfam: PF00454; P13_P14_kinase; 1.
 DR PROSITE: PS00915; P13_4_KINASE_1; UNKNOWN_1.
 DR PROSITE: PS0290; P13_4_KINASE_3; 1.
 KW Kinase.
 SQ SEQUENCE 1889 AA; 207787 MW; ED5828BD3A89557 CRC64;

Query Match 42.3%; Score 66; DB 10; Length 1889;
 Best local similarity 1.2%; Pred. No. 4.5e+02;
 Matches 22; Conservative 1; Mismatches 8; Indels 1773; Gaps 6;

QY 1 SRHXNS-----7
 DB 45 SRHLHSLALRLPAGAGAGSSIPAPLAFASHAFSPSPQSFAPAFSLRLPL 104
 QY 8 -----7
 DB 105 LAAPTCPLSSALSALMALADVDAPASAPLANAFLSMAAASPPTLLPADAPVASRL 164
 QY 8 -----7
 DB 165 LEFASSEAPPKAKGGEYAGENGVRVYQKFEFEVELEKREKVFRLIVHVLGE 224
 QY 8 -----7
 DB 225 GGEISDNVAVKRNAAKQVNSLSEFLKIRKRWREGAQLTRINTKLCCOAAVVLYR 284
 QY 8 -----7
 DB 285 SVSAMVDYSKSKDMLQOTLAFLEATKSCILSSMRKICEBELFCULANVKSCHIE 344
 QY 8 -----7
 DB 345 FGWKDALVDPTLNRLAAYRENDYEEEDGKEKAVPWRMLVRLALCLCKWE 404
 QY 8 -----7
 DB 405 VDMILPLFTEHLEEGDASSPSLRLRLDAISRVAQLCEFEKRESIVLMTSTYDKAK 464
 QY 8 -----7
 DB 465 AVGSAENNTVSPATEERETELPAGFLVATNLTSTKRSYDHRRLSLCSDVLAESK 524
 QY 8 -----7
 DB 525 SGSGADLMGPPLPAVAEICSDPDVSTVEPSLLKFLRLMFTYVLEGLAPRIOSQTPA 584
 QY 8 -----7
 DB 585 KPYSTLSLNTWESLSATLAVSGPYMNSKMCVAVQRIAGTPPLVYSSVKLEDELEIN 644
 QY 8 -----7
 DB 645 ALNPPSRKGNSEKAVGQRTLSAALSRYEVAAMTISGVKATYLLAVAFLEILRES 704
 QY 8 -----11
 DB 705 CNGGILSANTLAKNSAFCVPEYLLTTRNLPAVSQCLTAVVHAFVLSNEDRID 764
 QY 12 -----11
 DB 765 IGGADIRESVISVHACFLIKMSQDENVRDVSVKLLQLEKKEPQVLAHSSCVDLILI 824
 QY 12 -----11
 DB 825 SVNELTSGVSDPANVAVRSLYOKIAREMTLSALSTAPCTTGLIQRCKPSCAQRS 884
 QY 12 -----16
 DB 885 QHTADVVLISLEIRICSGKNDMGITANVAVPMASAAAASGAKKRAPITTEVSTAVY 944
 QY 17 -----16

DB 945 TATVKNHAGEIAGNRRLRSSKAGANTGNSPLCKQSAQPNQSEDFVLSRFVLLQDPVY 1004
 QY 17 -----16
 DB 1005 TAEKQIDNTVFRETCQSQTALLLDHWSDSRANLDGSQLRLLCMPAYICTPDAMET 1064
 QY 17 -----16
 DB 1065 GIFTWTWVSAPSLGPLYLAELVDAMLTIDYKCGLPASDNNYCGPDALNPHLISGEF 1124
 QY 17 -----16
 DB 1125 EAPKDPVEALLAHRLNLGFFIDREFVYRHDSIEQLLLGRMLQSTMSKANHFSHPRA 1184
 QY 17 -----22
 DB 1185 TGFETAMLGLKFCSCQSOSNLQCKMGLQLEDEYRAALGFAYAEFTESQKSPA 1244
 QY 23 -----22
 DB 1245 QREAGVSIFVHCLQRENSGSADSAKSGRGEFNMLDQHPVGSVDNYTTRERK 1304
 QY 23 -----22
 DB 1305 OLLIMLSQNEADRLVYMAQPIHTKQATFRKISSDKWIDHARTAFVDPRIAFSMIRF 1364
 QY 23 -----22
 DB 1365 PTNSALSSBITQVQTHLELRTIPEALPFTTPRAVDENSILQOLPHMAPCSYQALE 1424
 QY 23 -----22
 DB 1425 FLTPYKGRVMAVYLVLETPPEYTFEHPOLYOSLRIDDNLVGEYLLGAARSNL 1484
 QY 23 -----22
 DB 1485 FHLILWNLQGEVDDEPKFAAPKATFHSLLPAVREKIYDGETPBARDFEREFEFF 1544
 QY 23 -----26
 DB 1545 DKVTSISGVLPFLPEERBAGIKRELEKTPVGGDLVLPATNKKVRQYVDSGIPLOSA 1604
 QY 27 -----26
 DB 1605 AKVPIMTFENVVDRDGDPRDVPKQACIFKVGDDCHQDVIALQVIALDIFQAVGLNLYL 1664
 QY 27 -----29
 DB 1665 FRYGVLPFGPERGILEVYNTFSRRQKGETYDGLLEIFQDDYGVGSPFEAREMEMI 1724
 QY 30 -----29
 DB 1725 SSAGYAVASILLQPDNRINGILLFDSHRLVHIDFGFLEISPGNGGESAHRLSHEM 1784
 QY 30 -----29
 DB 1785 TQLLDPSGTNKSDDTWNQFLRLCVKGYLAGRRHMGITTVVLAVDGSLPCSRNGEPTANL 1844
 QY 30 --RF 31
 DB 1845 KRKF 1848

RESULT 10
 ID 042181
 AC 042181
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE PKD1 protein.
 GN PKD1.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97449170; PubMed=9285785;
 RA Sandford R., Sgotto B., Aparicio S., Brenner S., Vaudin M., Wilson R.,
 RA Chisoe S., Pepin K., Bateman A., Chotia C., Hughes J., Harris P.;
 RT Comparative analysis of the polycystic kidney disease 1 (PKD1) gene
 RT reveals an integral membrane glycoprotein with multiple evolutionary
 RT conserved domains.
 RL Hum. Mol. Genet. 6:1483-1489(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Vaudin M.;
 RT The sequence of Fugu rubripes 259c6.
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Washu;
 RT A Fugu rubripes Genome Sequencing Project.
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF013614; AAB86683.1; -.
 DR HSSP: P98161; 1B4R.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR001024; Lipoxigenase_LH2.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR002859; PKD/REJ-like.
 DR InterPro: IPR000434; PKD_L.
 DR InterPro: IPR000203; PKD_cys_rich.
 DR InterPro: IPR000601; PKD_domain.
 DR InterPro: IPR002889; WSC.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01825; GPs_1.
 DR Pfam: PF00059; lectin_c_1.
 DR Pfam: PF01463; LRCT_1.
 DR Pfam: PF00801; PKD_15.
 DR Pfam: PF01477; PLAF_1.
 DR Pfam: PF02010; REJ_1.
 DR Pfam: PF01822; WSC_1.
 DR PRINTS: PR00500; POLYCYSTIN1.
 DR SMART: SM0034; CLECT_1.
 DR SMART: SM00303; GPs_1.
 DR SMART: SM00308; LH2_1.
 DR SMART: SM00082; LRCT_1.
 DR SMART: SM00089; PKD_13.
 DR SMART: SM00321; WSC_1.
 DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
 DR PROSITE: PS50093; PKD_9.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 SO SEQUENCE 4578 AA; 504585 MW; 62AA07B61C4FC7A CRC64;

Query Match 42.3%; Score 66; DB 13; Length 4578;
 Best Local Similarity 0.68; Pred. NO. 1.9e+03;
 Matches 17; Conservative 6; Mismatches 8; Indels 2950; Gaps 4;

QY 1 SRXHXSM----- 9
 Db 466 ARKTHSERGVMLGLSDVNSAGLHWNGSEAGEGEGGLAPSSVARGNLVSLDQRSQ 525
 QY 10 -----XRPD----- 14
 Db 526 TSSHCNARAYVQYNPOVVPDGVFAMGLAVFPSSHLLQATRAMITTPPPHGGVEV 585
 QY 15 ----- 14

Db 586 LILPALSFRVAGRLSLSEFVTRLSQTHLEQIYRRCRYPGCGPACSPMALCVPSSES 645
 QY 15 ----- 14
 Db 646 RSADPPSPCKLEQCPYQNCPLPSRPC078SCINCQGH078EARMPLTYLOYEVLFS 705
 QY 15 ----- 14
 Db 706 LTAGRAHVLLQDQLEDLVSRGVIALQHDGAPASLINC0SPHSIMROPVVALNLSEW 765
 QY 15 ----- 14
 Db 766 FGINTFPVPDNLADGADPLDPDELKALVEGTRSGWLENAVCPTRVLYVQSETRLR 825
 QY 15 ----- 14
 Db 826 GAELSAGLPQPGVYLVLTSAKQSDVVSASGLQVLPDLTVLHSPQNGTLEFSNRT 885
 QY 15 ----- 14
 Db 886 RLLLRVQSRVETKAVYRTSNSSATFNPYCPQEFSSRLCHPSTSGAEEVAEPMYAVLD 945
 QY 15 ----- 14
 Db 946 LMLKTEHTGPVQVELKARNNTASLTVLVYLESLRGVYQPHRRVIMESVSTTA 1005
 QY 15 ----- 14
 Db 1006 SYLBSGNPTFKMTVDKRPFTYNTLVNVIYQHADVYKLTATAMNHSTLLEDNVTDR 1065
 QY 15 ----- 14
 Db 1066 LQPMNTLVYKGVDPVPOGSGTQTLTSLVLDMAVPATVWMSFGDGYEQFHKPPYSSL 1125
 QY 15 ----- 14
 Db 1126 TCEPSPYQVLLSNNTYIYSQGIYTAVVSNNRENISRSINKSVSILTRVDIQTEPS 1185
 QY 15 ----- 14
 Db 1186 LILAGTADFEAHPLPSPYGLHEWNPFGSALLQGRVATFEKSGPFNVCSVNTTIS 1245
 QY 15 ----- 14
 Db 1246 SRAVCTEMFAYEIESLAESSPTLHSATTVAHLSSGNNTYTPSMGDKSTYLSSEP 1305
 QY 15 ----- 14
 Db 1306 TVSHKXANDGRFTVNTLTAMNAVSGWTIIPVEFVFOVAGIEPSCGVEHALVNRARVS 1365
 QY 15 ----- 14
 Db 1366 GNASSYLEMSPGDSPNETHHNPGLSHTYRNGCTHPLSLITTSGVSKANFVQVCVP 1425
 QY 15 ----- 14
 Db 1426 VLTKISVTAEKSHFAVGEKIQOQARAEPPENNTYEMDGGEDVLPAPARVNTYNNP 1485
 QY 15 ----- 14
 Db 1486 GCYVTVAAASNNVSNMTGISIEVLTPGCAVVEQHNQTHYNNLTGLGVPYFSAFSSASNV 1545
 QY 15 ----- 14
 Db 1546 SYLMNFGHNLKGNLTATFNTPGHNNITLTAAVRVGNKTTVPAAVLASVGLTINS 1605
 QY 15 ----- 14
 Db 1606 LVAVPLNTSVHPEAHNDEDSVRSNILLDCHCTPLFRNNTHFTYTRSGVTENIITYAEND 1665
 QY 15 ----- 14
 Db 1666 VGAAQASTILFVQRELEGQLIVETAEGGGGVQELDMCYFETNRYVREHAGKEGTNLT 1725


```

OY 15 -----INPA----- 18
Db 1726 FTWVILKNEPDNISIFNLGTAKVNPBPCDIFLQANLIGOVNRTIYFLEPARN 1785
OY 19 ----- 18
Db 1786 VOLOISDPVAVNALINNTVLNMESSNLQYRFEVNCDDLQNSKSMSSHFTSAGOKOVTV 1845
OY 19 ----- 18
Db 1846 KVENESSEVSELSVQVQVIGLKIISADATBQNYFPPTDVSCLQGBLSTGTIVTSMWL 1905
OY 19 ----- 18
Db 1906 IDGSKMOKKACVTFPKPKYAVATLMTNDVSGKVVSRFPVQBRIFSELKASKXIAAI 1965
OY 19 ----- 18
Db 1966 NEKVEPSTISLVAGTRVNLSTLSISRDATVVLQPNQTYHTSRVDTIVVNLTAVNEVSKR 2025
OY 19 ----- 18
Db 2026 RNLIENVEPYRGLSIODSCAALAVGEKKEVAVNVTGKPVHFLMTFDLHILHKEHEMCK 2085
OY 19 ----- 18
Db 2086 EVSTAEBSGLITYLGAINALHONITROMAQNVLMADVLAIPODTFVKNKVTANVL 2145
OY 19 ----- 18
Db 2146 VTPKSNFMDCIMIFGDYSAPRHTNNTVGHETHPGHYRQVQVNCNLVSNVSAHAETNVS 2205
OY 19 ----- 18
Db 2206 VLECEBEVQVQAPRLAIVRSQPTLYEASVDLKGCLRGAQVLMQLLSAPSCNDPCHA 2265
OY 19 ----- 18
Db 2266 SGRVNGATRSFPVLPFAEDVDQRLQLSLPKVLAAGNYTLVSLXYENVPLKKAQCLQ 2325
OY 19 ----- 18
Db 2326 GVMARLAPIIBGTYRWNSQTDLQLSAESQYDPNNDPNQSLHYNHWCQSTSKPER 2385
OY 19 ----- 18
Db 2386 CSTLSFGISGKPVGLISGSELENGIETFEKLSIGKEMPESTGTQVLVOSGHIPWYL 2445
OY 19 ----- 18
Db 2446 ECVCCKAOSLYEVSQNSYVTLKGTCSQCQFHMGRMSAVTLQNDTLVLDSSFTTSGGM 2505
OY 19 ----- 18
Db 2506 NLVLKQGVLIHGSYNFTLHVTDSDLDGEGASITLHNNPDPGGBCHLRGGETQENG 2565
OY 19 ----- 18
Db 2566 DKEDVRIETLLDRVQSCSGSDIGVSETPILYSLVNRCDYCEBEFYVRCSSPEH 2625
OY 19 ----- 18
Db 2626 SATLPFSSAQHRVSVFTVEDHGAALAKTKIKVLPDPPEYSSILPHLSDLIDS 2685
OY 19 ----- 18
Db 2686 KLKLLDQGFQVRELISLVLYLNEYBQTRSVRSVRYRGVRRVNNIRALJALD 2745
OY 19 ----- 18
Db 2746 LITVNDIQOTSAAAOCTAVSRFPICEECONSTLNLESMLETLQDTQGIPTPEIAD 2805

```

```

OY 19 ----- 18
Db 2806 NILINMGDLHOVSQTSOSYNQPPYVDSPPSFIOKDKGGLDPSPPLEPHLVAANA 2865
OY 19 ----- 18
Db 2866 YSLSVLMLILHARVLINEPVLKGAETATCKLADPOLCYHGNNSPGGFTLHETLK 2925
OY 19 ----- 18
Db 2926 ANILFSSRLRALFSLHQVESNPPFPVYVNTVSTEVASMEFTEGTHIPISGLD 2985
OY 19 ----- 18
Db 2986 SLATVAVNNGSNGAENEPGAGTGGVPPAGA VNISSCDSVLYRSMGNTNRAGLFLVL 3045
OY 19 ----- 18
Db 3046 NTSLODVLQDKREAEELISITAVLSHEKPRFNCTDRKRITLGMTRGHDLHKKYTF 3105
OY 19 ----- 18
Db 3106 FLSPELYDTTLDYFIVNSTACGSGGACVRLVGVFASLCQYFSESEKQMTDGNVPLA 3165
OY 19 ----- 18
Db 3166 ETNARVACRHLTAFAGLFVPANALSTVPERSGTPSLVLYLVCLVGLMSYVMAAI 3225
OY 19 ----- 18
Db 3226 LHKLDLDRRAVVLGCGDGLFKYEQVYKWSKACTAHVGISLYGRRSGRHL 3285
OY 19 ----- 25
Db 3286 DSRGSRARALDIFHATDTSLGAVKAKIMDNKGLSPAMLQVLYVADLQGTSSSYFL 3345
OY 26 ----- 25
Db 3346 VEEMLSVNDEKTDGVEIYEASSEAMLLQPLRLRYELQALCESHMLSLFORPPSP 3405
OY 26 ----- 31
Db 3406 FTRLQATCCALLILLIANTLWTSIVVDVRYSPRVSRRF 3446

```

RESULT 11
010466 PRELIMINARY: PRT; 26926 AA.

AC 010466: PRELIMINARY: PRT; 26926 AA.

DT 01-NOV-1996 (Tremblere, 01, Created)

DT 01-NOV-1996 (Tremblere, 01, Last sequence update)

DE 01-JUN-2002 (Tremblere, 21, Last annotation update)

DE titin, heart isoform N2-B (EC 2.7.1.1-) (Connectin).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HEART;

RX MEDLINE=96026330; PubMed=7569978;

RA Labeit S., Kolmer B.;

RT "Titin: giant proteins in charge of muscle ultrastructure and elasticity."

RL Science 270:293-296(1995).

RN [2]

RP SEQUENCE OF 22277-25376 FROM N.A.

RX MEDLINE=92258380; PubMed=1582406;

RA Labeit S., Gautel M., Lakey A., Trinick J.;

RT "Towards a molecular understanding of titin."

RL EMBO J. 11:1711-1716(1992).

RN [3]

RP SEQUENCE OF 1976-2014 FROM N.A.

Label: S;
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE-95331314; PubMed-7607248;
 RA Gautel M., Castiglione-Morelli M.A., Pfuhi M., Motta A., Pastore A.;
 RT "A calmodulin-binding sequence in the C-terminus of human cardiac
 titin kinase."
 RL Eur. J. Biochem. 230:752-759(1995).
 CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
 ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
 SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
 PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
 DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
 ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
 N2-B.
 CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
 CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
 KINASES.
 CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
 IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
 DOMAINS.
 CC EXBL: X64698; CA45939.1; -
 DR EXBL: X83270; CA58243.1; -
 DR EXBL: X64697; CA45938.1; -
 DR EXBL: X90568; CA462188.1; -
 DR EXBL: X64699; CA45940.1; -
 DR HSP: P56276; 1TLK
 DR Interpro: IPR000282; Cytok_receptor_2.
 DR Interpro: IPR000719; Euk_kinase.
 DR Interpro: IPR000577; FGGY_kin.
 DR Interpro: IPR003962; Fm11_repeat.
 DR Interpro: IPR003961; FN_III.
 DR Interpro: IPR001092; Huh_baslc.
 DR Interpro: IPR003598; Ig_c2.
 DR Interpro: IPR003600; Ig_c2.
 DR Interpro: IPR003006; Ig_MHC.
 DR Interpro: IPR000129; Peptidase_S24.
 DR Interpro: IPR002016; Peroxidase.
 DR Interpro: IPR002290; Ser_thr_kinase.
 DR Interpro: IPR001245; Tyr_kinase.
 DR Pfam: PF00041; fn3; 132.
 DR Pfam: PF00047; Ig_91.
 DR Pfam: PF00069; kinase; 1.
 DR PRINTS: PRO0014; FNTYPE11.
 DR PRINTS: PRO0726; LEXASERPTASE.
 DR Prodom: PD000001; Euk_kinase; 1.
 DR SMART: SM00060; FN3; 127.
 DR SMART: SM00408; IGC2; 23.
 DR SMART: SM00410; IG_1like; 79.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
 KW serine/threonine-protein kinase; Alternative splicing; Repeat;
 KW Immunoglobulin domain; Phosphorylation.
 FT Domain 1370 1389 4 X 5 AA TANDEM REPEATS OF R-N-S-P-A.
 FT DOMAIN 4429 4614 GLU/LYS/PRO/VAL-RICH.
 FT DOMAIN 24731 25070 CATALYTIC.
 FT DOMAIN 25030 25056 CALMODULIN-BINDING.
 FT MOD_RES 1372 1372 PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 1372 1372 PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 1382 1382 PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 1387 1387 PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 26171 26171 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26171 26171 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26184 26184 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26184 26184 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26190 26190 PHOSPHORYLATION (POTENTIAL).

FT CONFLICT 22277 22277 T -> P (IN REF. 2).
 FT CONFLICT 22449 22449 E -> G (IN REF. 2).
 FT CONFLICT 22454 22454 T -> Q (IN REF. 2).
 FT CONFLICT 23324 23324 S -> L (IN REF. 2).
 SQ SEQUENCE 26926 AA; 2993428 MW; D5EECD3254F5523 CRC64;
 Query Match 42.3%; Score 66; DB 4; Length 26926;
 Best Local Similarity 0.1%; Pred. No. 3; 4e-04;
 Matches 22; Conservative 3; Mismatches 6; Indels 24391; Gaps 7;
 1 SRKH-----
 Db 2450 NTHASDSEGPLLVGRVETNCNLSVEKIKIRGLDLCTETQNVVFEVLSHGIDVL 2518
 QY 6-----
 Db 2519 WAFNDKEIPSSKYKIFAHGRIYKLVINMKMDGKTYFAGENMTSGKLTVGATSK 2578
 QY 6-----
 Db 2579 PLIDQTVASQDAVECEVANDPSKGEMLRDKHLPLTNTIRSEOGHRRLLIAATKLD 2638
 QY 6-----
 Db 2639 DIGETTYKAVNSKTSAKLKEAVKIKTKLTLTETQDAVYVELTIPNKGYQAIKNG 2698
 QY 6-----
 Db 2699 VVLESNEKALSVKGYISLRINKCAIVDESYYGRLGRLGASARLAVEIVKIKRPDY 2758
 QY 6-----
 Db 2759 TALENATVAEVSVDHPVPKPKSVKIKPSDKHLVSEKVKIKMLQNTSPSDAGEY 2818
 QY 6-----
 Db 2819 TAVGQLDCKALFVELLHITTKNIEVEPKTASFECEVSHFVPSMKLNGVEIENS 2878
 QY 6-----
 Db 2879 EKKIYVQCKLQLIMNTSTEDSAEYFVCGNDVSATLVTPIYIMTSMKDIKAEKD 2938
 QY 6-----
 Db 2939 TTFEVTVAEGISYKWLKNGVEIKSTDCQNRKTKLTHSINTRVHFGDAADVTFFAGK 2998
 QY 6-----
 Db 2999 ATSTATLYEARHIEFRKHIDIVLEKRRAMECEVSEPDITVQMKDDDELITDRIK 3058
 QY 6-----
 Db 3059 IQEKIVARHLIPSTRMSDAGKTYVAGGNSTAKLFEGRDVRIRSIKVEOVIEKORA 3118
 QY 6-----
 Db 3119 VVEFEVNEEDVDAMWKDIEINFOVERHKYVERRIHMFISRTQSDAGEYFVAGK 3178
 QY 6-----
 Db 3179 NRSVTLVYNAPPEQVQLQPVTVOSGPARFCAMISGRPKISYKKEQLSTGFK 3238
 QY 6-----
 Db 3239 CKELHDSQETVLLLEAFEDAAVYTCEAKNDYVATTSASLSVPEVSPDQEPVYP 3298
 QY 6-----
 Db 3299 PAITPLQDTVTSQGPAPFCQVSGTDLKVYSKDKIKPSRFRAQFDITQLEIA 3358
 QY 6-----
 Db 3359 EAVPEDEGTYTFVANNVAGVYSTANLSLEAPESILHERIDQIEKMKNEFSSFLSABE 3418

QY 6 ----- 5
Db 3419 EGLHSAELQJLSEINTELLSESPVYPPKEDSEKEGTGPIFKVSNADISMGDVATLSV 3478
QY 6 ----- 5
Db 3479 TVIGIPKPKIOWFENGVLTPSADYKVFDCDDHSLIILFTKLEDESEYTCMASNDYKGT 3538
QY 6 ----- 5
Db 3539 ICSAVLKNKSGCHKDTESESAAKSLKIGCPPHFLKELKPIKCAQGLAIFETV 3598
QY 6 ----- 5
Db 3599 VGEPAFVYTWKRNKQCTSVYTYIINPNPNSGTFIVNDPQREDSGLYICKAENMLGESV 3658
QY 6 ----- 5
Db 3659 CAABELLVLEBDTMDTRCKAKSTPEAPDEFPOTPLKPAVEALDSBOETATFVKDTILK 3718
QY 6 ----- 5
Db 3719 AALITEENQOLSYEHIAKANELSQLPLAQELQJLEQDKLPESUREFLCINGSIHQ 3778
QY 6 ----- 5
Db 3779 PLKPSNLOLOIYOSOKTFSKGIIMPEEPFOAVLSDTEKIFPSAMSTEQINSILTEP 3838
QY 6 ----- 5
Db 3839 LKTLAPBEGNPOSSIEPPNHSYLSVABEVLSEKENTVSDINREORVTLQKQOASAL 3898
QY 6 ----- 5
Db 3899 ILSQSLAEGVESIQSPDVAMISQVTEPLVPSHSECTEGKILIESANPLENAGDSAVV 3958
QY 6 ----- 5
Db 3959 IEBGKSLRFLALBEKOVLLKEHSDNVNMPPOITTESKREPAIKVQFVQGRDLSRE 4018
QY 6 ----- 5
Db 4019 SLTSGIPEORNLKIQICBALQAAVASEQPLFSEMLNIEKVEVEAVNITQEPHIMC 4078
QY 6 ----- 5
Db 4079 MYLVTSKSVTEEVYITIEDVDPQMANLKMELDALCAIIEBIDILTAEGPRIQGAKT 4138
QY 6 ----- 5
Db 4139 SIOBENDSPSGQKVEPIPEPVESEKYLISTEVSYFNQSRVKYLDATPVTKGVASAVV 4198
QY 6 ----- 5
Db 4199 SDEKODESLPSEKESESSSESGTEEVAVTKIQEABGLIKEDCPMILHTPLVDTVSEGD 4258
QY 6 ----- 5
Db 4259 IYVLLTSTNAKAVNTPFENKLVPSDEKFKJLQDQNTYTLVLDKNTVEDHOGYCEALN 4318
QY 6 ----- 5
Db 4319 DSGKATSAKLIVVKAAPVIRKIEPLEVALGHLAKFTCEIQASAPNNAFQFKAGRELY 4378
QY 6 ----- 5
Db 4379 ESDKCSIRSSKYSLSLEILRTQVDCGEYTCASNEYGSVCTAILTVTPGGEKKVRL 4438
QY 6 ----- 5
Db 4439 LPERKPEKEEVVLSVLAKRBEERKVEPKLEKVKPDAVPEPPPKPVEVEVPTV 4498

QY 6 ----- 5
Db 4499 KREKPIPEPTKYPEIKPAIPLAPBPKPPEAEVKTIRPPVEPEPTIABVTVPVGK 4558
QY 6 ----- 5
Db 4559 KAKAKAPKEBAKPKGPIKGVPKTSPLEABERKLRPSGGGKPEDEAPFTYQKAVPL 4618
QY 6 ----- 5
Db 4619 KFYKEIKDILIESEFVGSALFECVSPSTAITTWMKQSNIRSPKHFADKDRKL 4678
QY 6 ----- 5
Db 4679 HLDVQDSAGETCVLRLGNKEKSTAKLVYELPVRFVKTLEEVYVYKQPLYSCE 4738
QY 6 ----- 5
Db 4739 LMKRDVVRKDKIVENKGRIVRGVIGIMBALNINDADTDAGTYTVENANNLSCS 4798
QY 6 ----- 5
Db 4799 SCYKVVETIRDLVYKPIRQHVKPKGTALFACDIANKDPNINWFKGYDEIPAEPRDKTEI 4858
QY 6 ----- 5
Db 4859 LKDCNHLVYKINAMPEDIAEYAVELEBKRPYAKILIGREVELLKPIEDVTYKESAS 4918
QY 6 ----- 5
Db 4919 FPAISEADIPQWIKIGELLRSPCEIKAGGRFLIHKVKIDQAGEVLYQALNAT 4978
QY 6 ----- 5
Db 4979 TALLVYKEIJDFAVPLADVYPERQARFECVLTREANVTMSKPDILIKSSDKFDITAD 5038
QY 6 ----- 5
Db 5039 GRKHVLVINDSQFDEGVYTAVEGKTSARLVYIGIRLKFMSPLEDQTVAGEFATFVC 5098
QY 6 ----- 5
Db 5099 ELSHEKNAVYKNDAKLHTSFTVLISSDGTHKLEKEVTLDDISQIKAOVKELSTAQ 5158
QY 6 ----- 5
Db 5159 LKVLADPYFTVKLHDKTAVEKDEITLCEVSKQVPVPMFKDGEIYVSPKSTKADGLR 5218
QY 6 ----- 5
Db 5219 RILKIKKADLNDGEGYCDGDKTKANVYVBARLIEYKBLYGVEVGEIATHEIELS 5278
QY 6 ----- 5
Db 5279 EPDVGQKLLKQGPLASPDCEIIEDGKHLILLHNCQLGTEGVSPQANAKSAAMLKV 5338
QY 6 ----- 5
Db 5339 KELPLIFITPLSDVAVFEBKDEAKFCEBVSREPKFRMLKGTQITSGDRFELINDGTRHS 5398
QY 6 ----- 5
Db 5399 NVIKSAPEDEAKVYFEADKRTSGKLLITBIRLKLPLADVYAKERESAVFVELSHD 5458
QY 6 ----- 5
Db 5459 NLRVKNFNQDLRHTTRSVSNQDEGKTHSITFKDLSTDTSQIVELAKMSSEAKLTVLE 5518
QY 6 ----- 5
Db 5519 GDPVFTGRLQDYTGKDEVLIOCEISKADAPVWFKDGEIRPSKNAVINTDKKRMILI 5578
QY 6 ----- 5

```

Db 5579 LKRLKSDIGQYTCDCGTDKTSKGLIEDREIKLVRLHSVEVETAFETIEISDDI 5638
QY 6 ----- 5
Db 5639 HAWMKLNGEALLQTPDCEIKEEGKIHSLVLHNCRLDQTGGVDFQANVKSAAHLRVKPRV 5698
QY 6 ----- 5
Db 5699 IGLIRPLKDVTVAGETATEDGELSTYDIPVWYLGKLEPSSDKVVPSEGVHTITLIR 5758
QY 6 ----- 5
Db 5759 DVKLEDAGEVQLLANKPETHANLFVKBPVEFTKPLEDQVEGATVLECEVSRENAY 5818
QY 6 ----- 5
Db 5819 KMFKNGTEILKSKRYEIVADGRVKLYIHDCPBDITTYCDANFXTSCNLVVPVHE 5878
QY 6 ----- 5
Db 5879 FLRLPLDQVKEKARECELSRENAKVKFMDGALIKKGYDIISKAVRLVYNKC 5938
QY 6 ----- 5
Db 5939 LLDEAEYSCVETARTSGMLTVLEBAVFTKLANIEVSEITDIKLVCYSKPGAEVW 5998
QY 6 ----- 5
Db 5999 YKGBEIIETGRYELTEGRKRLIVIONAHLEDAGNTNCRPLSSRTDGKVKVHELAAFTI 6058
QY 6 ----- 5
Db 6059 SKPQNLILBGEKAFVCSISESPVQWKRDKLTLESQKYDYIADGKRVLVVKNATL 6118
QY 6 ----- 5
Db 6119 QDMGYVAVGAAAPAAHLIVIEKRIYVPLKDTRVKQOEVENCEVNTEGANAKVFN 6178
QY 6 ----- 5
Db 6179 BEATDSSKTIILQKOLVYTLRIRDAHDDQANYNSLTHNGENVKSANLIVEEDLR 6238
QY 6 ----- 5
Db 6239 IVEPLKDIETHEKSVTFWCKVNRNLVTLKWTNGEEVFPDNRVSYVDKXKHLTIKDC 6298
QY 6 ----- 5
Db 6299 GFPEGEYIVTAGODKSVALLIIEAPLEVEHLEDQVTEFDVAVFSQLSREKANVK 6358
QY 6 ----- 5
Db 6359 YRNGEIEGKKYKFEKSGSIHRLIINKDRLDDCEVACSGVEDKRSARLVEELPEI 6418
QY 6 ----- 5
Db 6419 RPPQDILEAPGADVFLAELNKDVEYQMLRNNVVYQGDHOMSEKIHRLQICDIKP 6478
QY 6 ----- 5
Db 6479 RDQGYRIANDKEARAKLELAAPKITADQDLVVDGKFLINVPYDAYPKAAEWFK 6538
QY 6 ----- 5
Db 6539 ENEPSTKITDTTAQTSFRILEAKKGDKGRKIYVQNHKKBGSFIMLXIVDPGPVRN 6598
QY 6 ----- 5
Db 6599 LEVTEIFQGEVSLAWEPLTDGSKIIIGYVERRDIKRTVWLADRAESCEFTVTSIQK 6658
QY 6 ----- 5

```

```

Db 6659 GGEVYLFRYSANRKYGTGPEVEDNPNVEARSKYDVPGLNVTITVDNRFVSLTWEPPE 6718
QY 6 ----- 5
Db 6719 YDGAELITNYVIELDKDTSIRMDTAMTVRAEDLSATVDYEGQESFRVAQRNIGVK 6778
QY 6 ----- 5
Db 6779 PSAATPEVAVADPIERSPVNLTSDDQTSQSVQLKMEPPLKDGSPILGYIIEECERK 6838
QY 6 ----- 5
Db 6839 DNMIRCNMLVPELTLYKYTGLEKGNKLYIRVAENKAGVDSPEILGPLJADDAVEPTM 6898
QY 6 ----- 5
Db 6899 DLSAFKDGLEVIVPNPITLIVPSTGYPRPTATWCSDGVLEFGDVRVNMKTLASVLEVIS 6958
QY 6 ----- 5
Db 6959 PSEBSDKGIYTLKLENVXTISGEIDVNIAPRSAPKELFGDITPKDSVHLWEPPDDG 7018
QY 6 ----- 5
Db 7019 GSPITGYVVEKREVSRTWTYKWDVFTDLEFTVPLVOGKEYLFFVCARNKCGEPAYV 7078
QY 6 ----- 5
Db 7079 DEPVNKPATVDPDPENKVRDRTANSITLWDPKNDGSGRIKGYIVFCRPGSDKV 7138
QY 6 ----- 5
Db 7139 ACGEPAETEMEYTGLEBKMYAYRYKTLNRQASRSPHTEIQAVIDQEAPELFDVK 7198
QY 6 ----- 5
Db 7199 LLAGLTVKAGITIELPATYTGPERKITWKADHILKQDKRTIENVRKSTVITVDSKR 7258
QY 6 ----- 5
Db 7259 SDITGYIIEAVNVCGRATAVEVNVLDKPGPAARDITDVNESCLITWNPDRDGSKI 7318
QY 6 ----- 5
Db 7319 TNYVERRATDSEVWHKLSITVKNFKATKLIPNKEYLFRVAENMGAGEVQASPT 7378
QY 6 ----- 5
Db 7379 AKYQDPGPPPTLPSDITKDAVTLTNGEPDDGSPITGYVERLDPDTOKVRCKNM 7438
QY 6 ----- 5
Db 7439 PVKDTYVRVGLNKKRYFRVLAENLAGPKPSKSTEBILIKDPIIDPPWPGKPTVADV 7498
QY 6 ----- 5
Db 7499 GKTSVRLNMTKPBHDGAKTESTVIMLKTGDEWVRAVEGPTTQHLPLGMBQESVF 7558
QY 6 ----- 5
Db 7559 RYRAVNKAGESPESDPVLCREKLYPPSPRMLEVINITKNATADAKVPEKDGSPIT 7618
QY 6 ----- 5
Db 7619 TNYIVEKDVRRKQWIVDTTVNDTKCTVPLTEGSLYFRVAENAGISDYTELEDSV 7678
QY 6 ----- 5
Db 7679 LANDFTTGPYPALAVDVTKRHVDLKMEPPKNDGGRPIQRYVIEKKERLSTRVYKAK 7738
QY 6 ----- 5
Db 7739 TAGPDCNFRVTDVIEGTEVQFQVRAENAGVGHSPETELISTEDPTSPSPPLDLHVTD 7798

```

```

OY 6 ----- 5
Db 7799 AGRHIALAMPKPKNGSPILIGYHMECPVGI EKWRVNSRPIDUKFEVERGVDPKE 7858
OY 6 ----- 5
Db 7859 YVLFRAVNAIGVSESEISENVAKDPDCKPTIDLETHDIYIEGKLSIPVPEFAVAV 7918
OY 6 ----- 5
Db 7919 PTVSMHDKGEVKASDLTKMNDHISAHLEVPKSVRADAGIYTTLENKLSGSAIASINVK 7978
OY 6 ----- 5
Db 7979 VIGLPGCKDIKASDITKSSCKLTWEPPEPDGTPILHLYLEREAGRTYIIPVMSGENK 8038
OY 6 ----- 5
Db 8039 LSWTVKDLIPNGEYFRVAVKAVNGGSEYIEKNPVIAQDPKPPDVPVYEVHNPTAEA 8098
OY 6 ----- 5
Db 8099 MTITWKDPLVYGSKIMGYIIEKIANGEERWKRKCNELHPLITATKGLKEGKEQFRVR 8158
OY 6 ----- 5
Db 8159 AENRAGISEPSRATPTKAVDPIDAPKVIILKTSLEVRGDEIALDASISGSPYPTTWIK 8218
OY 6 ----- 5
Db 8219 DENVIPEIEIKRAAPLVRRRKGEVQEEEPFLPLTORLSDNSKKGESQLRVRDSLRPD 8278
OY 6 ----- 5
Db 8279 HGLYMKVENDGIANAPCTSVLDTPGPPINVFEDIRKTSVLCWEPPLDDGSGEITN 8338
OY 6 ----- 5
Db 8339 YTLEKDKTPDSEMIYVITLHCKYKSYTKLIEGKEYIFRVAENRFGDPPCVSKPLV 8398
OY 6 ----- 5
Db 8399 AKDPFGPADPKPIVEDYTSNSMLVKMNEPKDNGSPILGYMLEKREVNSTHMSRVNKSJ 8458
OY 6 ----- 5
Db 8459 IALKANVDGLLEGITYFRVCAENAGPKESPSPDPKTAHDPISPGPPILRVYDTSS 8518
OY 6 ----- 5
Db 8519 TIELEMPFPAFNGGEIYCIYFDKQLVGNKKSRCETEMIKYKQYTVAREINEGADYKLR 8578
OY 6 ----- 5
Db 8579 VSAVNAGGEPGEIOPVTVAEQEPFAVELDVSKGIIQIMAGKTLRIPAVVTGRPVPT 8638
OY 6 ----- 5
Db 8639 KYWTKEEGELDKDRVIDVNGTKSELLIIDLAKDGRVYITATNSCGSKFAARAVEVD 8698
OY 6 ----- 5
Db 8699 VPGPVLDLKPVTNNRKMCLLANSDDPDGSEITGFIIEKDAKHTHROPLETESKCD 8758
OY 6 ----- 5
Db 8759 IIGLLEGQYKERVAKNKGCGPVEIGPILAVDPLGPTSPERLTYTEROSTITLDM 8818
OY 6 ----- 5
Db 8819 KEPRSNGSGPIGYIIEKRRHOKPDFERVKNKLCPTTSLVENLDEHOMYFRVAVNFI 8878

```

```

OY 6 ----- 5
Db 8879 GSESEPLPLNVVYIODEVPPITKLRLSVSGDTIKVAGEPVHLPADVGLPKRIENSKN 8938
OY 6 ----- 5
Db 8939 ETVIERPDALQITKEEVSSEAKTELSPKAVREDKGYTVTASNRUGSVFRNVHEVY 8998
OY 6 ----- 5
Db 8999 DRESPRNLAVDIKAESCYLTVADAPLDNGGSEITHYIDKDRSKKAEMEETNTAVE 9058
OY 6 ----- 5
Db 9059 KRYGKWLIPNGQYEEFRVAVANKYGISDECKSDKVVYIOPYRLPGPKPVLAFTKSGM 9118
OY 6 ----- 5
Db 9119 LVSMTPLDNGSPITGYWLEREGSPYSHVSHAPITKGLAKGVFRNVRLLGVRVQ 9178
OY 6 ----- 5
Db 9179 FRAMAINAGIGPSPSPDPEVAGDPIFPPGPPSCPEYKDKTKSSIISLGMPRANGSGP 9238
OY 6 ----- 5
Db 9239 IKGIVEMQEGTDMKRVNEPDKLITTCQCVPMILKLRKFRVAKAVNAGESEPSDT 9298
OY 6 ----- 5
Db 9299 TGEIATDIOEBEYFIDIGADCLYCKAGSQRIPAVIKGNPPTKSSWEDGKAKKMK 9358
OY 6 ----- 5
Db 9359 DGVHIDPEDAQETAENSVIIIPCKRSHGKYSITAKNKAQOKTANCRKRVADVPGP 9418
OY 6 ----- 5
Db 9419 KDLKVSIDITRSGRLSMKPPDDGDRJIKGYIEKRTIDGKAWTVNPDGSTEYVVDL 9478
OY 6 ----- 5
Db 9479 LSEQYFRVRAENRFGIGPVEIQTITANDPIYPPDPTIKLIGLITKNVHLKMKP 9538
OY 6 ----- 5
Db 9539 KNDGSPYTHYIECLANDPTGKKAWRQCNKRDVELOFTVEDLYVSGEYFRVAVN 9598
OY 6 ----- 5
Db 9599 AAGVSRPATVGPCCQORPDPMSIDLEFMEVEGETNVIANKIKGVPTLTFAPR 9658
OY 6 ----- 5
Db 9659 KRPDKPEVLYDHYNKLVDCTCLVLPQSRSDTGLYITAVNNLCTASKBMLNVLG 9718
OY 6 ----- 5
Db 9719 RGPVGPPIKESVSADQMLSHPPKDDGSKITNVYIEKREANRRTWVHSEPECT 9778
OY 6 ----- 5
Db 9779 YTIPLLEGHEVVRIMAKNNGYIGEPDSEPTARNLFVPGADPVTSSVTRNSMTV 9838
OY 6 ----- 5
Db 9839 NWEPEYDGSPPYGYWLEMDTTSKRKRVNRDPIKAMTIGSVYKTYGLIGSDYQFRV 9898
OY 6 ----- 5
Db 9899 YAINAGVPASLPSDPATADPIAPGPPPKVTWTKSSADLEWSPPLKDGSKVIGY 9958
OY 6 ----- 5

```

Db 9959 IVEYKGEKREKGNKDEYKRTKLVYTGKESGAFYKFRSANNAGIGEPGEVTVIEM 10018
 QY 6 ----- 5
 Db 10019 KDRIVSPDLQLDASVDRIVVAGVIRIIVYSGRPPPTVTNNMNERLPEQATIEETTA 10078
 QY 6 ----- 5
 Db 10079 ISSSMVIEKNCORSHOGVYSILANNEAGERKKTIIIVDLVDPGEVGPPELAIHNLINESCKI 10138
 QY 6 ----- 5
 Db 10139 TWTSPEDDGGSPITNTYIEKRESDRAMPVYTVTRQNAVQGLIQKAVPFRIMAENS 10198
 QY 6 ----- 5
 Db 10199 IGMGPVESEALVIREPTVPERPEDLEVEKYTKNTVLTNPMPKDGSEIINTVLES 10258
 QY 6 ----- 5
 Db 10259 RLIGTEFKHVTDNLISKRYTVKLGKESDTEYKRSAVNIYOGKPSFCTKFTICKDEL 10318
 QY 6 ----- 5
 Db 10319 APPTLHDFRDKLTIRVGEAFALGRYSGRPKPVSWFKDEADVLEDDRIHKTIPATLA 10378
 QY 6 ----- 5
 Db 10379 LEKIKARSDSGKYCVVENSTGSRKFCQVNVVDHPGPVSGVEDETKDYMTISWKP 10438
 QY 6 ----- 5
 Db 10439 PLDDGSKINTYIIEKEVEGKDWMPYTSASAKTCTKVSLLGCKOYIPRIHAENLXGIS 10498
 QY 6 ----- 5
 Db 10499 DPLVDSKAKADRRVPDADPDQPIVETKDSALVTNNKPDGSKPTNTYIEKRETMKR 10558
 QY 6 ----- 5
 Db 10559 RMAVATDPIHPYIKFVRPDLLEGCOYEFRVSAENEIGIDPPSPSKPVFAKDPJAKPSP 10618
 QY 6 ----- 5
 Db 10619 PVPNEADITTCNSVDLTWOPRHDGSKILGYIYEQKVGDEMRANHTPESCPETKYK 10678
 QY 6 ----- 5
 Db 10679 VTGLRDSQTYKFRVLAVNAGESDPAHVPEPVLVKDRLEPPELILANMARERHIVGDT 10738
 QY 6 ----- 5
 Db 10739 LRLSAILKGVFPFVTKKEDRDPYKADIVTVPGSKLEIRNAHEDGGLSLTVENPA 10798
 QY 6 ----- 5
 Db 10799 GSKTVSVAVLADKPPRDLEVEIRKDSCLTFWKEPLDDGGSVITNTYVVERDVASAQ 10858
 QY 6 ----- 5
 Db 10859 WSPISATSKRSHFAKHLNEGNOYLFVNAENOGRGPFVETPPKIRALDPLHPPGPPKD 10918
 QY 6 ----- 5
 Db 10919 LHHVDVDTKTEVSLWNKPRDGGSPITGYIYEOBEGTODHIKKTYTNLECYVTGLQOG 10978
 QY 6 ----- 5
 Db 10979 KTYRFRVAKENIVGIGLBDTTIPTEQCKIAPPSEVLQVKLIEGLVYKAGTVPAPAIR 11038
 QY 6 ----- 5

Db 11039 GVPVPTAKTIDISEIKDEHYETVETDNFSSVLTIKNCILRHDGTGEQYIIVSNAGSKTVA 11098
 QY 6 ----- 5
 Db 11099 VHLTVLDVPGPEPGBINILDTPEHMTISWOPPKDDGSPVINTIVEKODTRKDTMGVVS 11158
 QY 6 ----- 5
 Db 11159 SGSSKTKLKIPLHOKGCEYFRVRAENKIGVPLDSPTVAKHKFSPSPGKPVYTDI 11218
 QY 6 ----- 5
 Db 11219 TENATVSWTLPKSDGSPIIIGYMERREYTKWVRNKKTEPIADLKRRVTGLYEGNTYEF 11278
 QY 6 ----- 5
 Db 11279 RVFAENLAGLSPSSDPITACRPITKPPPINPKLKDQKRETAADIWMTPLSDGSPIT 11338
 QY 6 ----- 5
 Db 11339 LGYVEEQKPGTAQMNRIKDELJROCAFVYVGLIEGNEYFRILKAANIYEGSEPRELAE 11398
 QY 6 ----- 5
 Db 11399 SYIANDLHPPEVELDYTCRDVITVRGQITRILARVGRPEPDTWTKEGKVLVREKRV 11458
 QY 6 ----- 5
 Db 11459 DLIODLPRVLOIKFENVRADHKYIISAKNSGHAQSAIYVLDPRPQCNIAVTNTYK 11518
 QY 6 ----- 5
 Db 11519 ENCTISWENPLDNGSEITNFIYERKPNOKGSIASDVYERLIKANILANNEYFRVC 11578
 QY 6 ----- 5
 Db 11579 AENKVGSPITETKPIILINIDRGPENLHIAKCKTFYLLKRRPDYDGGSPNLSTY 11638
 QY 6 ----- 5
 Db 11639 HYERRLKGSDDKERVHKSIRKEIHYMDRCVENQIYEFYQTKNKGSDMWKTEEVYVK 11698
 QY 6 ----- 5
 Db 11699 EDLQKPVLDLKLGVLTVRAGDTIRLEAGVNGKPEPEVAMTKNDATDILTRSPVRIDPR 11758
 QY 6 ----- 5
 Db 11759 AOSKFSILKAKRSDGKTVTATNTAGSFVAIATVNVLDKGPVRANKIYDVSDDCTV 11818
 QY 6 ----- 5
 Db 11819 CNDPPEDDGCEIQNYILEKCEIKRWMSYSAVLTPTGYTRILEGNEYFRVRAENK 11878
 QY 6 ----- 5
 Db 11879 IGTPTESKPYIAKTYDKPGRPPPEYTVKSKEMTVVNNPPEYDGGSLTGYFLEKK 11938
 QY 6 ----- 5
 Db 11939 EKHSFRVVPVNSAIPERKKYONLLPDHEYOFRAENEIGIGESLSRPRVAKDPTE 11998
 QY 6 ----- 5
 Db 11999 PPGPTNFRVYDTTKHSITLGMGKPYDGGAPLIGYVEMRPIADASPDGCKRQNAAL 12058
 QY 6 ----- 5
 Db 12059 QLVKKEFTVSLDENQYEFRCANOQVIGRPAELKPAIKKEIPEPEIDLDASMRKL 12118
 QY 6 ----- 5
 Db 12119 VIVRAGCPILRLFAIVGRBAPKVTNRKVGIDNVTRKGQVLDVDMALVITPNSINDSGK 12178

QY 6 ----- 5
Db 12179 YSLTLVNPAGEKAVFNVEHLDIPGVSDLKVSDYKTSCHVSWAPENDGGSQYTHYIV 12238
QY 6 ----- 5
Db 12239 EKREADKRWSTVTPVVKATSFHVTNLVPGNEHYFRVTAVNEYGHGVPDVKPVLASDP 12298
QY 6 ----- 5
Db 12299 LSEPPPKLETEMKNGSATLAWLPPLNDGAKIDGYLISYREERPADRWETEXSVKD 12358
QY 6 ----- 5
Db 12359 LSLVVTGLKGGKKYFRVARNVAVSLPREAGVTEAKEDLLPPKILMPEQITTKAGKK 12418
QY 6 ----- 5
Db 12419 LKIEAHVYGNKPHPTCKKKKGEDEVYTSHTLAVHKADSSILLIKDYTRKDSGYSLVAMN 12478
QY 6 ----- 5
Db 12479 SSGTDIQIKVYVWDAPGPPPPEDISLDADACSLSHIPLJEDGGSNTINYIVEKCVS 12538
QY 6 ----- 5
Db 12539 KEDWATALASVTKTSCRVGKLLPGQETJFRVRAENRFGISEPLISPKVAVQPPGVPSEP 12598
QY 6 ----- 5
Db 12599 KNAHYTKVKKDCIFVAMDBDSDGSPILIGYLIERKERSLLMVANDTVRSTYPCAG 12658
QY 6 ----- 5
Db 12659 LVEGLEYSFRIYALKAGSSPPSPLEYVTARMPVDPGKPEVIDVTKSTVSLIMARPKH 12718
QY 6 ----- 5
Db 12719 DQSKTIIGYTVACKLPDGRVRCNTAPHOIQOEETATGLEKAQYFRALARTAVNIS 12778
QY 6 ----- 5
Db 12779 PPSPPDPVTLAENVPPRIDSVAMKSLIVKAGTNCIDATVGRKMPVSMKKDGTL 12838
QY 6 ----- 5
Db 12839 LKPAGIKMAQORNLCTLELFSVNRKDSGYITTAENSGSKSATIKLVLDPGPASV 12898
QY 6 ----- 5
Db 12899 KINKMTSDRAMLSEPPLEDGSEITNTYVDKRETSRPNMAQVSATVPITSCSVEKLEIG 12958
QY 6 ----- 5
Db 12959 HEYQFRIKAEKNGYGDVTEPALAKNPYDPGRCDPPVSNITKDMATVSMKPPADG 13018
QY 6 ----- 5
Db 13019 GSPITGYLLEKRETOAVNMKTVNRKPIITERLKATGLQGETEYEFRTAINKAGRGKPSD 13078
QY 6 ----- 5
Db 13079 ASKAAVADPOVPPAPPAFPKVIYDITRSSVSLSMGKPAIDGSGPILIGYLVYKRAQSDNW 13138
QY 6 ----- 5
Db 13139 YRCNLPQNLQKTRFEVTLGIMEDQYQFRYAVANKIGYSDPQVPOKHYPKDILIPBEGH 13198
QY 6 ----- 5
Db 13199 DADLAKTLILRAGVTWRLVYPVKGPPKPTWSPKNVNLRORTIGLIDIXSTDEFTLRCEN 13258

QY 6 ----- 5
Db 13259 VNKYDAGYILLLENSCGKKEFTYIVKVLDPGPPINVTYKELSKDSAYTWEPPIIDG 13318
QY 6 ----- 5
Db 13319 SPIINYVQKDBARRKSWSTYTTECSKTSFRRVPMLEBGSYFFRFAENEYIGDPQETR 13378
QY 6 ----- 5
Db 13379 DAYKASQTPGPVVDLKVRVSVKSSCSIGMKKPHSDGSRIGYVYDFLTENKQORVMS 13438
QY 6 ----- 5
Db 13439 LSLQYSAKDLEGEYTFRVSANENEGPSEITIVARBDVYAPDLGLDLCYLAK 13498
QY 6 ----- 5
Db 13499 ENSNFKLPIKGPAPBSVSKGEDPLANDTRVSVESAVENTTLIVYDQKSDAGKTYI 13558
QY 6 ----- 5
Db 13559 TLKNVAGTKECTISIKVYGNKGIPTGPIKFDVTAEMATLKNAPKDDGSEITNTILEK 13618
QY 6 ----- 5
Db 13619 RDSVNNKWTGASAVQKTFEVRBLHEGMEYTFRVSANENYGVGEGLSKSEPIVARRHFDV 13678
QY 6 ----- 5
Db 13679 PDAPPPNIDVNRHDSVSLTWDPKKTGSGPITGYMLEFRNSLLMKRANKTPIRMRDP 13738
QY 6 ----- 5
Db 13739 KYTGLTEGEYFRVMAINLAGVKPSLSEPPVVALDPIDPPGKPEYINITRNSVTLIRT 13798
QY 6 ----- 5
Db 13799 EPKYDGHKILGYLVEKRDLPSSKMAKNHVNPECAFVTLVLEGGKYEFRIKANKTAG 13858
QY 6 ----- 5
Db 13859 AISAPESTETIICKDEYAPVLDPTIYDGLTIKAGDVIYNAISLICKPLPKSSWSK 13918
QY 6 ----- 5
Db 13919 AGKDIRSDITQITSTPSSMLTIKYATRKDAGEYITITATNPSTKVENHYKTVLDVPGP 13978
QY 6 ----- 5
Db 13979 PGVYISNVSAEKATLMTPLLEDGSPKSYLIERKETSRLMTVSEDIOSGRHVAJK 14038
QY 6 ----- 5
Db 14039 LIQNEHYIFRVSANHNGKGPVQSEPYKAVDRPGPGPEKPEVSNVTKNATVSMKRP 14098
QY 6 ----- 5
Db 14099 VDDGSEITGYHVRREKSLRWFAIKTPVSDLRKVTLQSGSTYEFRVSANENRAGIG 14158
QY 6 ----- 5
Db 14159 PPSEASDVLMKDAAPPGBPSPNPHVTDTTKSASILAMGRPHYDGLGITGYVYHQKYG 14218
QY 6 ----- 5
Db 14219 DEAWIKDTGTALRLIQFVYVPLDQTKENYFRISALINDAGVGPVAVDPVEYIEREMAD 14278
QY 6 ----- 5
Db 14279 FELDDELRLVVRAGLSIRIFVPIKGRPAPEVTWMDINILKRNANIENTESFTILLIP 14338
QY 6 ----- 5

Db 14339 EGNRYDTGKFWMTLENPAGKSGFVNVRVLDTPGVLNLBPTDITKDSVTLJHDLPLIDJ 14398
 QY 6 ----- 5
 Db 14399 GSRITTYIVKREKATRSYSTATTCKHCICYKTGLSECEYFVRMAENFYIGBPTET 14458
 QY 6 ----- 5
 Db 14459 TEVPKASEAPSPDLSLIMDITKSTYSLSLAWPKRKHOGSRTTYGYVLEAQRKSGDQWTHII 14518
 QY 6 ----- 5
 Db 14519 TVKGLCEVANTLEGEETPOVAVNSAGRSAPRESHPVTKQTMJPELDNGIYOKLY 14578
 QY 6 ----- 5
 Db 14579 IANAGNIKVEIPVLGRPKFTVTKKQDQLKOTQRYNFTTASTILINBECVSDSGP 14638
 QY 6 ----- 5
 Db 14639 YPLARNIAGEVDVITIOYHDLPGPTGPIKFDVSSDEVTSKDPENDGVPISNTY 14698
 QY 6 ----- 5
 Db 14699 VEMKQDSTTWVBLATVIRTPYKAPRLITGLEVOFRKAQNRXVGPGITSAMIVANYP 14758
 QY 6 ----- 5
 Db 14759 FKVPGPPTQOVTAVNTDSMTISHHPLSDGSPILGIVHERKERNGILMOTYSKALVPG 14818
 QY 6 ----- 5
 Db 14819 NIFSSGLTQIAIEFVIAENNAKSKPSKPSEPMALDIPDPGKVPVNLTRHTVTL 14878
 QY 6 ----- 5
 Db 14879 KWANPEYTGFKITSYIYERKDLJPNGRMLKANSNILENFIYSGITEDAAYEVRVIAKN 14938
 QY 6 ----- 5
 Db 14939 SKDKELEGTAKLEIKIADPSTVLNKRDSRDSGATTLATNPGRAKHIFNVKVLDRP 15058
 QY 6 ----- 5
 Db 15059 GPPGPIAVTEVTSEKCVLSNPPPLDDGAKIDHYIVQKETSRLANTNVASEVOYTKLK 15118
 QY 6 ----- 5
 Db 15119 VTKLLKNEIYFRMAVNNKGVGEPLSEPVLANPYGPPDPKPNPVTITTKDSMVVCW 15178
 QY 6 ----- 5
 Db 15179 GHPSDGSSEIINIVYERBKAGORIKCNKKTLLDLRYVSGLTBEGHEYFRIMAENNA 15238
 QY 6 ----- 5
 Db 15239 GISAPSPSPRYKACDVFVKPGPNRVLDTSRSSISLANNKPIYDGGSEITGYVWEIA 15298
 QY 6 ----- 5
 Db 15299 LPREDENOIYTPRAGLKATSTITIGLTENQEKIRITAMNSEGLGEPALVPGIPKADRM 15358
 QY 6 ----- 5
 Db 15359 LPPEILDADLRKVITIRACCTLRFPVPIKGRPDEVKARHDSGLDKASIESASSYTL 15418
 QY 6 ----- 5

Db 15419 LIVGNVNRFDGSKYIIVFENSNGSKSAFVNVRVLDTPGPQDLKVEKTKISTVLTWDP 15478
 QY 6 ----- 5
 Db 15479 LIDGSKTKNTIYVKRESIRKAYSTVATNCHKISWKVDQLOGCSYIFRVLAENEYIGL 15538
 QY 6 ----- 5
 Db 15539 PAETASVKASERPLPGKITLMDVTRNSVSLSMKPEHDGSRLLGYIVEMQTKSDKM 15598
 QY 6 ----- 5
 Db 15599 ATCATVYKTEATITGLIGEEYSFRVSAQNEKISDPRQLSVPIADLVIPAFKLLFN 15658
 QY 6 ----- 5
 Db 15659 TTFVLAGEDLKVDPFIFGRPTPAVTHKDNVPLKOTTRVNAESTENSLLIKDACRDV 15718
 QY 6 ----- 5
 Db 15719 GHVVKLINSAGEALFTLVNVLDRPGPTGPKMDEVLTADSLTSMGPFRKDGSSINN 15778
 QY 6 ----- 5
 Db 15779 YIVKRDSTITWQIVSATVARTIIRACRLTKGCEYOFRIAENRYKSTYIASEPTVAQ 15838
 QY 6 ----- 5
 Db 15839 YPFKVPGPPTVYVTLSSDSMEVQWNEPISDGSNVIYGHLEKERNLSILMWLNTKPI 15898
 QY 6 ----- 5
 Db 15899 POTKFTTGLEGEVEYEFVSAENIVGIGKPSKSECVARDPCDPGRPAIIVTRNSV 15958
 QY 6 ----- 5
 Db 15959 TLOMKRPYDGSKITTYIYVEKKELPDGRMAKASTNIIDHFEVTGLVEDHRYEFVIA 16018
 QY 6 ----- 5
 Db 16019 RNAGVSESESTGATARDEVDPFRISMDPKYDITVIHAGESFKVDADITGKPIPTI 16078
 QY 6 ----- 5
 Db 16079 QWIKGQELSNARLEIKSTDPATSLSVKAVRDSGNYILKKNVAGERSTVNVKVLID 16138
 QY 6 ----- 5
 Db 16139 RPPPEGPVYIGVTAEKCTIANKRPPLODGSDIINYVERRETSRLVWTVVDANVOILS 16198
 QY 6 ----- 5
 Db 16199 CKVTKLBNETFRIMAVNKYVGEPLSEBPVAKNPFVVDAPKAEVTVTKDSMIV 16258
 QY 6 ----- 5
 Db 16259 VNERPASDGSSEILGYVELEKRDKEGIRMTCHKRLLIGELRLRWTLINHDYEFVSAEN 16318
 QY 6 ----- 5
 Db 16319 AAGLSBPSPSAVOKACDPIYKPPPNPKVIDITRSSVFLSMKPIYDGCCEIOGYIVE 16378
 QY 6 ----- 5
 Db 16379 KCDVANGETMCTPPTGINKTNIYEKLEHEYNFRICAINKAGVGBHADVPPIVEE 16438
 QY 6 ----- 5
 Db 16439 KLEAPDIDLELRKTIINRAGGSLRFPVPIGRPTPEVKMGKVDGEIRDAAILDYTSF 16498
 QY 6 ----- 5
 Db 16499 TSLVTLDNVNRKDSGKITLLENSSGTSKAFVTVVRVLDTPSPPVNLKVEITKDSVSIWE 16558


```

OY 6 ----- 5
Db 16559 PLLDGGSKINXIVKREARVYAAVVTNCHRNMRKLDQLOEGSYFRVTAENEXG 16618
OY 6 ----- 5
Db 16619 GLPAQTADPIVAEVPQPKITVDVTRNSVLSMTKPEHGGSKTIQYIVEMQAKHSE 16678
OY 6 ----- 5
Db 16679 KMSCARVKSQAVITNLTQGEYLFRVAANENKGRSDPRSLAIPYAKDLVLEPDYKA 16738
OY 6 ----- 5
Db 16739 FSSYSVOVGODLKKMVPISGRPKPTIWTIKDGLPKQOTRINTVSDILTLISIKETHKD 16798
OY 6 ----- 5
Db 16799 DGGQYGLIVANVVGQKTASIEITVLKRPDPKGFVKFDVSAESITLSMRPPLYTGSCQI 16858
OY 6 ----- 5
Db 16859 TNYIVKRDITTYVMDVVSATVANTILKWKTKGTGYOFRIFAENRYGGSFALESDDIV 16918
OY 6 ----- 5
Db 16919 AQYFKEPGPPTFPANAISKDSKVIOHMEPVNNGSPVIGYHLEKERNSLMTKYNKT 16978
OY 6 ----- 5
Db 16979 LIHDTQKQNLBEGIEFEFRVYVAMNIVGAKASKNSECYVADPCDPGTPPEIMYKRN 17038
OY 6 ----- 5
Db 17039 EITLQMTKPYVDGSMITGYIVKRDLPDGRMKASFTNVLEQFVSGLEDOYEFYR 17098
OY 6 ----- 5
Db 17099 IAKNAAGAIKSPSDSTGPIAKDEVLPRIKMDKFRDITIVANGETFRLADVHGKPLP 17158
OY 6 ----- 5
Db 17159 TLEWLRGDKELIESARCEIKNTDFKALLIVKDALRIDGQYILRASNVAGSKSPVAVKY 17218
OY 6 ----- 5
Db 17219 LDRPPEGPVOYTVTSEKSLTWSPLODGGSDISHYVERKETSRLAMTVVASEVVT 17278
OY 6 ----- 5
Db 17279 NSLKVTKLLEGNEVFRIMAVNKYVGEPLDESAPVLMKNPFVLPGPCKSLVETNIAKDSW 17338
OY 6 ----- 5
Db 17339 TVCMNRPDSGSEIIGYIVKRDGIRWIKCNKRITDLRLAVTGLFEDHEFEYVSA 17398
OY 6 ----- 5
Db 17399 ENRAGGEPSPATVYKADPVEKPGPTNAHIVDTKNSITLWGRPIYDGGSEIIGYV 17458
OY 6 ----- 5
Db 17459 VELCKADEBMQIVTQGTGLRVTREFELSKLTHEQYKIRCAKVKGLGSAVSVPQTVAP 17518
OY 6 ----- 5
Db 17519 EDRLAEPELDLDELKRGIVVRAGSARIRHIFKGRPMPELTVSREGEETDKVQIEKV 17578
OY 6 ----- 5
Db 17579 NTYQSLDNCDRNDAGKIYIKLENSSGSKSATVTVKVLDTGPPONLAVKEVAKDSATLV 17638

```

```

OY 6 ----- 5
Db 17639 WEPPIIDGAKVKNYIUKRESTRKAYANVSSKSTSPKVENTEGALITYFRMAENEF 17698
OY 6 ----- 5
Db 17699 GVGPEVETDAKRAEPSPGKVITLDVQSISLIMKEKPEHGGSRVIGYVEMQPKG 17758
OY 6 ----- 5
Db 17759 TEKWSIYAESKVCNAVVTGLSSGOEYOFKVKAYNEKGDPRVLGVPVIAKDLTTOPSLK 17818
OY 6 ----- 5
Db 17819 LPENTYSIQAGEDLKLEIPVIGRPRPNISWVKDGEPLKQOTRVNVEBATSTVLHKGN 17878
OY 6 ----- 5
Db 17879 KDDGKYTVATNSACTATENSVILKRPVGVPRDEVSADEFVVISMEPRAYTGCC 17938
OY 6 ----- 5
Db 17939 QISNTIYKRDITTTTWHVVSATVANTTIKIKTKGTGYOFRIFAENRYGKSADLSKA 17998
OY 6 ----- 5
Db 17999 VIVQYFKEPGPPTFPVTSISKQMLVOMHEPVNDGTRKIIYHLEQKENSILMYKLN 18058
OY 6 ----- 5
Db 18059 KTIPIQTKRKTGLDGLYEFRVSAENIVGIGKSKVSECFVANDPCDPGRPEALVIT 18118
OY 6 ----- 5
Db 18119 RNNVTLKKKKPAVDGSKITGYIVKRDLPDGRMKASTNVLETFYVSGLEVDQRYEF 18178
OY 6 ----- 5
Db 18179 RVIAARNAAGNFSBPSDSGATARDEIDAPNASLPKYKDVIVHAGETFEVLEADIRGKP 18238
OY 6 ----- 5
Db 18239 IPDVWMSKDKLEETAPAMEIKSTIOKTTLVKDCINTDGGQYILKLSNMGTRKSIDIT 18298
OY 6 ----- 5
Db 18299 VKYLDKPSBEGPIKVTGYIAKCYLIAMNPLQDDGANISHYIIKRETSRLSTQVSTE 18358
OY 6 ----- 5
Db 18359 VOALNKKVTKLLPGNDYIFRVAVAVNKYIGEBLESQPVYACNPRPGRPGSTPEVSATIK 18418
OY 6 ----- 5
Db 18419 DSNVYTNAPVDDGGEIEIGYILKRDKEGVWTKCNKKTLDLRLAVTGLTGHSTYEF 18478
OY 12 ----- 11
Db 18479 VAAENAGVGESEPSVFYACDALYPPGPPSNPKVTJMSRSSVSLAMSKPIYOGGAPVK 18538
OY 12 ----- 11
Db 18539 GYVEVKEAAMDEWTCPPGLGQKOPTVYIKLNTENTYFRICAINSEGVGEPATLPGS 18598
OY 12 ----- 11
Db 18599 VVAQERIEPPELEADLRKVVYVLAASATLRLFTYIKGRPEVWKEAEBGLIDRQOIE 18658
OY 12 ----- 11
Db 18659 VTSFTMLVINDVTRFDSGRVNLTLNNSGSKTAFVAVRVLDSAPVNLRLREVKNDSV 18718
OY 12 ----- 11

```

Db 18719 TLSWEPPLIDGAKITNVIKRETTKRAVATITNNCTKTFRIENLOEGCSYFVRLAS 18778
QY 12 ----- 11
Db 18779 NEYIGLPAETTEPVKSEPLPPGKRVTLVDVTRNTATIKENKESDGSKITGYVEMQ 18838
QY 12 ----- 11
Db 18839 TKGEKSTCTOVKTLEATISGLTAGERYVRVAANVEKGRSDPRQJGVPIANDIEIKP 18898
QY 12 ----- 11
Db 18899 SVELPFETFNKAREOLKIDVPFKRPOATVNMKDCQTLKETTRVNVSSKTVLSLIR 18958
QY 12 ----- 11
Db 18959 EASKEDVGTLELVNSAGSITVPIITLVIDRPGPPGIRIDEVSCDSTITSMNPPEYDQ 19018
QY 12 ----- 11
Db 19019 GCQISNIVYKKEITSTWHLVSAVARTSLIKIVRLTGSYOFRCVCAENNYGKSSYSES 19078
QY 12 ----- 11
Db 19079 SAVVAEYFSPBPGGUPKVVHATKSTMLVTWQVPVNDGSGRYIGYHLETKERSIIMSK 19138
QY 12 ----- 11
Db 19139 ANKILIDTQYVSGLDEGLMTERVTAENTAGIGKCSCEPVAPADPCDPPQPEVTN 19198
QY 12 ----- 11
Db 19199 ITRKSVSLMKSKPHDGAKITGYIVERRELPGDKMLKNTNTIQEYFEVTELTEDQRY 19258
QY 12 ----- 11
Db 19259 EFRVAFARNADSVSEPSSTGPITVKDVEPPRVMDVKFQVIVKAGEVLKINADIAG 19318
QY 12 ----- 11
Db 19319 RPLPVISAKDGEIEERARFEIISTDNHITLITVKDCIRDTQGVLTLMVAGTRSAV 19378
QY 12 ----- 11
Db 19379 NCKVLDPGPPAGPLEINGLAKKCSLSGRPOEDGADIDYHKKHRETSNLAMTICEG 19438
QY 12 ----- 11
Db 19439 ELQMTSCHVTLKNGNEYIFRVGVNKYVGEPLSAIKALDPTVPSPPTSLEITSVT 19498
QY 12 ----- 11
Db 19499 KESMTLCSRPESDGSISGIIIRREKNSLRVVRNKKVYDLRVKSTGLRGCEY 19558
QY 12 ----- 11
Db 19559 RYLAENAGLSIPSETSPILNAEDVFLPSPSKPKIVDSGKTTITLAVAPLFDGAPI 19618
QY 12 ----- 11
Db 19619 TGYIVEYKKSDDTKTISIQSLRGTEYITISGLTTGAETVVRKSVKAVGASDPSDSDQ 19678
QY 12 ----- 11
Db 19679 IAKERBEPLPIDSEMAKITLVKAGASTMTVPFRGRFVNVLMKSDPTDLTRAYVDT 19738
QY 12 ----- 11
Db 19739 TDSRTSLTENANRNDGKITTLIQNLASAISLVLVVLDPGPPNTITVQDVTKESAV 19798
QY 12 ----- 11

Db 19799 LSWDVPENDGAPVKNYHLEKREASKKAVSVTNNCNRLSYEVNLOGAITYFVSGEN 19858
QY 12 ----- 11
Db 19859 EFGVGLPAETTKESVATTEKPPPEKIGVTSISKDSVSLTWLAPENDGSRIVNVYEALE 19918
QY 12 ----- 11
Db 19919 KGOKNWKCAVAKSTHHVYSGLRENSEYFVRFAENQAGLSDPRBELLPLVLIKQLEPPE 19978
QY 12 ----- 11
Db 19979 IDKKNFSTHYVVRAGSNLKVDPISGRKPLFKVTLSDGVPLKATMRENTETITANLITN 20038
QY 12 ----- 11
Db 20039 LKESVTADGREITTAANSQITKAFINIVLDRPGPTGPVVISDITEESTLKEPPK 20098
QY 12 ----- 11
Db 20099 YDGSQVNTVILLKRETSVAVTEVSATVAFMMKMLTGEYOFRIKAREFGISDH 20158
QY 12 ----- 11
Db 20159 IDSACVYKLPYTPGPPSTPVNTNTRESITVGWHEPVSNCGSAVGYHLEMKDRNSIL 20218
QY 12 ----- 11
Db 20219 WOKANKLIVRTTFKVTITISAGLIYERYAENAGVGRPSHPSEPVLAIDACPPRRVYR 20278
QY 12 ----- 11
Db 20279 IIDISKNVSLSWQOPAFDGSKITGYIVERRDLPGRWTKASFVNVEOPTISGLQON 20338
QY 12 ----- 11
Db 20339 SQYEFVAFARNAGVISNPSSEVGPITCIDSTGQVYIDLPLEYTEYVKTACTSVKLRAG 20398
QY 12 ----- 11
Db 20399 LSGKPATLEWYKDKELQTNALVCVENTDLASILIKDADRLNSGCELELRNMAASAS 20458
QY 12 ----- 11
Db 20459 ATRVQILDKPPGPPGPEIEKVTAEKITLMLRPPADDGAKITHYVERKETSRYVWSM 20518
QY 12 ----- 11
Db 20519 VSEHLECIITTKIITKNGEYIFRVAVNKYIGIGEPLESDSVAKNAFTVPPGPIPEVT 20578
QY 12 ----- 11
Db 20579 KITKNSWTVWSRPDIADGDSISGYFLEKDRKSLGWFVYLKETIRDTROQVGLTEND 20638
QY 12 ----- 11
Db 20639 YQYVCAVMAAGGPFSEPSFYKADPIDPPGPKIRIADSTKSSITLGMKSPYDGG 20698
QY 12 ----- 11
Db 20699 SAYTGYVEIRQGESEEMTVSTKGEVRTTEYVYVSNLKGVMYFENSAVNCAGGEPLE 20758
QY 12 ----- 11
Db 20759 MNEPVQAKDILEAPEIDIVALTYSIAKAGEDVQVILPFKGRPPTVYWRDEKNLIGSD 20818
QY 12 ----- 11
Db 20819 ARSIENTDSSLLITPQVTRNDTGKYYILLTENGVEGPFSSIVSVKVLDPACQMLQYK 20878
QY 12 ----- 11
Db 20879 HVSRGVTLLMDPPLIDGSPITINVIYKRDATKRTWVSVSHKCSSTFKLIDSEKTPF 20938

```

QY 12 ----- 11
Db 20939 FFRVLAENEIGIGEPETTEPVRKAEVAPAPIRDLNMDSTKTSVILSWTKPDGGSVLT 20998
QY 12 ----- 11
Db 20999 EYVERKKGGEOWSHAGISKEICEIVSOLKEOSVLEFRVFAKNENGSLDPYTTIGPIYVK 21058
QY 12 ----- 11
Db 21059 ELITTEVDLSDIPGAOVTVRIGHVHLELPYKGPCKPSISMUKDGLPKESFVRFSKT 21118
QY 12 ----- 11
Db 21119 ENK1TJLSIKNAKKEGKQYTVIIDNAVCRIAVPTVTLTGPPSKPGPIRDEIKADSVI 21178
QY 12 ----- 11
Db 21179 LSWDPEDNGGGEIYCSLEKRETSOTNNKMWSSVARTTEKVPNLVDAEYOFRVRAEN 21238
QY 12 ----- 11
Db 21239 RGVVSQPLVSSILVAKHQFRIBQPGPKPYLYNVTSDKMSLTWDAPYDGGSEVTGFHVEK 21298
QY 12 ----- 11
Db 21299 KERNSILMOKVNTSPISGREYRATGLVEGLDYOFRVYAENSAGLSSPSPSKFTLAVSPV 21358
QY 12 ----- 14
Db 21359 DPGFIDYIDYITREITLKNPPLRDGSKIVGYSTERKQGNERNVRCNFTVSECQTV 21418
QY 15 ----- 14
Db 21419 TGLSPEDRYEFRIANNAVGTISPPSQSGITMRDENVPPIVERGPEYFDGLIKSGES 21478
QY 15 ----- 14
Db 21479 LRITKALVOGRPVPRVTKWEOGYEIEKRMGETTNVLGTSLEVRATDRHGRVYVEAKN 21538
QY 15 ----- 14
Db 21539 ASGSAAKELKVAVODTPGKAVGPIRFTNITGEKMTLMDAPLNDGAPITHHYIEKRETS 21598
QY 15 ----- 14
Db 21599 KLAWALLIEDKCEAOSTYALIKLINGNEYOFVRSVANKFGVGRPLDSDPVVAQIQYTVDPAP 21658
QY 15 ----- 14
Db 21659 GIPESNITGNSITLWNAPESDGSEIQYTLERREKSTWVAVISKRPISETRFRVT 21718
QY 15 ----- 17
Db 21719 GLTEGNEYEFHWAERAGVGPASGISRLIKCREVPNPGPTVAVKVDTSKTVLEMS 21778
QY 18 ----- 17
Db 21779 KVPVGGMEIIGYIIEKCTDLGDWHAENAECVTRVTVTDLOGEYKRVASALINGAG 21838
QY 18 ----- 17
Db 21839 KGDSEVVTGKAVDRLAPELDIANFKQTHVVRAGASIRLFIAYQGRPIPTAVWSKPD 21898
QY 18 ----- 17
Db 21899 SMLSLRADIHDTDSFTLVENCNNDAGKYTLVEENSSKSLFTYKVLDPGPGPI 21958
QY 18 ----- 17
Db 21959 TKKDYTKGSATLMDAPILDGGARLHHYVEKREASRSRMOVISEKTRQIFKVNDAEG 22018

```

```

QY 18 ----- 17
Db 22019 VPYFRVASVNEGYGEPEYEMPEPIVATEQAPPRRLDVDTSSSAVLANKPDHGG 22078
QY 18 ----- 17
Db 22079 RITGYILEMROKGDLMWEAGHTKQLTFTVEALVEKTEYFRVAKNDAGSEPREAPSS 22138
QY 18 ----- 17
Db 22139 VIKRPOIEPTADLTGITNOLITCKAGSPFTIIVPISGRPAKVTWKLENRLETRDVS 22198
QY 18 ----- 17
Db 22199 ITTKDRTTLFVKDSMRDSDGRYFLENTAGVTFESYTVVVIQRPGVPTPIEVSSVA 22258
QY 18 ----- 17
Db 22259 ESCVLSMGEPRDGGTEITNTIVKRESGTTAMOLVNSVKTQIKVHLKMYEFSRV 22318
QY 18 ----- 17
Db 22319 SSENRFVSKPLESAPITAEHPVPPSAPTRPEYVYHSANAMSTRWSEPTHGSKITIGY 22378
QY 18 ----- 17
Db 22379 WYKKEERTIIMVXENKVPCLECNKVTGLVEGLEYOFRYALNAAGVSKASEASRPIMA 22438
QY 18 ----- 17
Db 22439 QNPVDAQRPREVTDVTRSTVSLINSAPAYDGSKVGVIIEKRPVSEVGDKMLKNTYI 22498
QY 18 ----- 17
Db 22499 VSDNFTVTLSEGDYEFRLAKNAGVLSKSESTGPTCRDBYAPRAELDARLHGD 22558
QY 18 ----- 17
Db 22559 LVTRAGSDLVIDAAGKPEPKIIMTKGKELDICERKVSQYTGKRAVAVIKFCDRSDS 22618
QY 18 ----- 17
Db 22619 GKTYILYKNAAGTAAVSVMVAVLDSPGCGKLTYSRVTOEKLWLSLPOEDGAEITHY 22678
QY 18 ----- 17
Db 22679 IVERRETSLMWYIYEGBCPTLSYVTRLIKNNXYIFRVAVANKYGPVYSESPYARN 22738
QY 18 ----- 17
Db 22739 SFTIPSPGPIPEVGTGKEHIIQWTKPESDGNESNLYLVKREKESLMTVRNNDYV 22798
QY 18 ----- 17
Db 22799 YDFRLKTYLSMEGDYOFRTAVNAAGNSEPSERSNF1SCREBSIYGPSPADRVVDTK 22858
QY 18 ----- 17
Db 22859 HSISLAWTKPAYDGTIDVGYVLEMQEKDIDOWRVHTNATINTEFTVPDLAKGKYSF 22918
QY 18 ----- 17
Db 22919 RVAAVNKGSESEYSAIEIEVERIEIPDLELADLKTVTIRAGASRLMVSVSGRPP 22978
QY 18 ----- 17
Db 22979 PVITWSOGIDLASRALIIDTSESILLIYDKVNRVYDAGKTTIRAEKNGSKASATVLYKY 23038
QY 18 ----- 17
Db 23039 DTPGPCSVKVEYSRDSVTITWEIPTIDGAPLNNIYVEKREAAHRAFKTVTTKSKTL 23098
QY 18 ----- 17

```

Db 23099 YRISGLVEGTMHFRVLPENITIGICEPCEISDAVLAYSEVPLVPAKLEVVDTKSPVTLAM 23158
 QY 18 ----- 17
 Db 23159 EKEPLYDGSRLTNGYLEACRAGTERMKKVTLKPVLEHTVSLNGEQYLEFRIRAOÑEK 23218
 QY 18 ----- 17
 Db 23219 GUSEPREVTAVVQDLRLPTIDSLTMPQKTIHVPAGRPELVLPFIAGRPPASWFA 23278
 QY 18 ----- 17
 Db 23279 GSKLRSESVYVETHTKVAKLFIRETTIRDTGEBTELEKNVGTSETIKVILIDKPGPP 23338
 QY 18 ----- 17
 Db 23339 TGPDKIDEIDATSITISWEPPELDGAPLSGYVVEQDRAHRCWLPVSESVTRSTFKETR 23398
 QY 18 ----- 17
 Db 23399 LTGCEVFRVATNFRGISTYQSEVIEGRSSIRIPGPETQIFDVSROGHTLWVPP 23458
 QY 18 ----- 17
 Db 23459 EDDGSGVGTGYIVERKEVRADRWVRKVPVMTVRYRSTGLEGEYEHRTVAINARSGS 23518
 QY 18 ----- 17
 Db 23519 KPSRSPPIVAMPDLPAPGKPRQNRVDTTRISVSLAMSVPEDEGSKVTGYLIMQKYD 23578
 QY 18 ----- 17
 Db 23579 QHEWTKCNTPTKIRETYTLHLPOGAEYRFRVLACNAGRGEPAEVPGVYKTEMLEPPD 23638
 QY 18 ----- 17
 Db 23639 YELDERYQEGIFVRQGVRLHTPIKGRPPICKWKTEGQDISRAMIATSETHELVAK 23698
 QY 18 ----- 17
 Db 23699 EADRGDSGTDLVLEKCGKAKVIAKYRVIGNSPBGPLEYDIOVRSVSKRRPADD 23758
 QY 18 ----- 22
 Db 23759 GGADILGVIERREVPKAAVTTIDSRVGTSLVWKGLENEYEHFRVSAENOGISPLK 23818
 QY 23 ----- 22
 Db 23819 SEEPVTPKTPLNPPPPSNPPVLDVTKSSVSLSMSRPKDDGSRVYGYIERKEKSTDK 23878
 QY 23 ----- 22
 Db 23879 VVRHNKQIOTTTWTVTGLVPAEYOFRIIAONDVGLSETSPASEPVVCKDPKPSQPG 23938
 QY 23 ----- 22
 Db 23939 ELEILISKDSVTLQWKEPECDGKEILAGVWEYRQSGSAMKSKNERIKDQFTIGL 23998
 QY 23 ----- 22
 Db 23999 LEATEYERVAFENETGLSRPRJAMSIKIKLISGEAPGIRKEMKVTTILGSAOLSCQ 24058
 QY 23 ----- 22
 Db 24059 IVGRPLPDIKMYRKGELIOSRKYMSSDGRHTLLVWTEEOBEDEGYVCLATNEVEVE 24118
 QY 23 ----- 22
 Db 24119 TSSKLLQATPOFHPGYPLKEKYGAVGSLRLHVMATIGRVPAMTFHQOKLLONSENI 24178
 QY 23 ----- 22

Db 24179 TIENTEHTHLVMKNVOKTHAGTKVQLSNVEGTDAIILDEIODKPRKPTGPVIEAL 24238
 QY 23 ----- 22
 Db 24239 LKNSAVISMRPPADGGSMITINVEEKEAKEGEMQJVSASISVTCRLVNLNENGY 24298
 QY 23 ----- 22
 Db 24299 FRVSAONTFGISDPLEVSSVVIKSPFEKPGAPGKPTITAVTKDSCVAMKPPASDGAK 24358
 QY 23 ----- 22
 Db 24359 IKNYTLERREKKQNKWISVTEBEIRETVFSYKULIEGLEFRVKENCENLGESEWSEISE 24418
 QY 23 ----- 22
 Db 24419 PITPSDVPIDAPHEKELRLNRYOSNATLVCKYVGHPRPIYKMYROGKEIADGLKY 24478
 QY 23 ----- 22
 Db 24479 RLOEFGGYHOLLIASVTDDBATVYQVRATNOGSGVSGIASLEVEPAKIHLPFTLEGMG 24538
 QY 23 ----- 22
 Db 24539 AVHALRGVSVSIKIPESGKPDVITWQODLIDNNGHYQVITRSTPSLVFNGVERKD 24598
 QY 23 ----- 25
 Db 24599 AGFTYVCAKNRFGIDQKTELDVADVDPDPRGVAVSDASRDSVNLTWTEPASOGSKITN 24658
 QY 26 ----- 25
 Db 24659 YIVERCAATTAERWLVRQARETRYVINLFGKTSYQFVIAENKFGLSKSPSEPTTK 24718
 QY 26 ----- 25
 Db 24719 EIKTAAMVDERVDETREVSMKASHSTKELYEKYMAIEDLGRGEGIVARCVETSKK 24778
 QY 26 ----- 25
 Db 24779 TYNAKEVVKGTQVIVAKKEISILINIAHRNLIHLHESFESMEIWMIFEEISGLDIFER 24838
 QY 26 ----- 25
 Db 24839 INTSAPELNEREIVSYHVQCEALOFLESHNIGHEDIRPENITYQTRSSITKIIIEGQA 24898
 QY 26 ----- 25
 Db 24899 ROLMGDNRLFLTAPEYAPVHQHDVSTATOMKSLCTLVYLLSGINFLAFTNOOI 24958
 QY 26 ----- 25
 Db 24959 IENINAAEYTFDEBAFKELISEAMDFVRLVYKERKSBMTASEALQHPWLKOKIERVSTK 25018
 QY 26 ----- 25
 Db 25019 VITRLAKHRYHTLIKKDLNMYSAARJSCGATRSQGVSAKYKVASIEIGVSGQIM 25078
 QY 26 ----- 25
 Db 25079 HAVSEGCHVKYVCKIENIDOSTQVWTFGVNQLENSEKEYTEIEYDGAIIIVKDIKLD 25138
 QY 26 ----- 25
 Db 25139 DETYRKYVNDYGEDSYAELEVKGVREYDYICRRTYKKIKRRDTNRLLERPPEFTLP 25198
 QY 26 ----- 25
 Db 25199 LYKTRAYGEENRFGVTTIVHPEPHVTKSGQKIKPGDNKKYTFESDKGLYQULTINSV 25258
 QY 26 ----- 25
 Db 25259 TTDDDAEYTVVARNKYGEDSCAKLTVTLHPEPTDSTLRPMFKRIIANAEQOGOSVCFE 25318

```

OY 26 ----- 25
Db 25319 IRVSGIPPLKMEKDGOPISLGPNIIEHGLDYALMIIDLPEDTGYRVATNTAG 25378
OY 26 ----- 25
Db 25379 STSQAHQVERLAKKOEKSKKEENHVOKOIDTLRAEILISTGESVPLQVANEAL 25438
OY 26 ----- 25
Db 25439 REAAVYPVAVSTKIVGEPFLIEIEKEERKLMFYDPEPRKYQTLIEEQRIKQFY 25498
OY 26 ----- 25
Db 25499 PMSDMKTKKIRDOYEMPGKIDRVQKRPKRILSMWEOFYVPLPRITDQYRPKWRIPK 25558
OY 26 ----- 25
Db 25559 LSQDDEIVPARRRTPSPYDYFRPRRRSLGDISDEILLPIDYLAKKTEEBRL 25618
OY 26 ----- 25
Db 25619 EEBELGFSASPPSPRPHELSLRYSPOAHVYETRNKFRYSTYHIPTKASTSY 25678
OY 26 ----- 25
Db 25679 AELREHAQAAVQPKORIMAREDELLRPVTTQHLSEYKSELDFNSKEEKSKKS 25738
OY 26 ----- 28
Db 25739 RROREVTEIEIEEYEISKHAQRESSASRLRRRRSLPTYIELMRPVSELINSRPO 25798
OY 29 ----- 28
Db 25799 PAEEYEDDTERRSPTPERTPRSPSPVSSRRSLRERSARFDIFGRYSKAAKTOKT 25858
OY 29 ----- 28
Db 25859 SERKEVLSQOPTLIDHARITLIRMSHRVPCQONTRFILINOSKPTAEVKYRNCVEIQ 25918
OY 29 ----- 28
Db 25919 ESSKHYTNTSGVLTLEILDCHTDGTYRACVCTNKGEASDVATLDVYGGDTTYASOR 25978
OY 29 ----- 28
Db 25979 RDEEVPBSVPELITRENAVAPSEFKTSEMASSSVREYKSOQTEETRESLSYHSASAE 26038
OY 29 ----- 28
Db 26039 MMSNALEKSLERKSTTRIKITLARIILTRPSMTVYBESARSCDTDEBPVTWVL 26098
OY 29 ----- 28
Db 26099 RKGQVISTARHQVTTTKYKSTFEISSVOASDEGNTSVVENSEGQAEFTLITQKARV 26158
OY 29 ----- 28
Db 26159 TEKAATSPPRVKSPEPRVKSPEAVKSPKRVKSPESHPRASPTETKPTPRKQVHLVPS 26218
OY 29 ----- 28
Db 26219 APPKITQFLAKASKRIALITCVESVSYRAKEVYTWYKDGKLENGHOFHYASDGYE 26278
OY 29 ----- 28
Db 26279 LKINNLTESDQGEYCEISGEGTSKTNLQFGQAFKSIHEKYSKISETKSKDQTTTEST 26338
OY 29 ----- 28
Db 26339 YNRKTEPKAPPEISSKPVIVTGLDPTVSSDVANFAVYKANGEPPTAIWTRDGAITOG 26398

```

```

OY 29 ----- 28
Db 26399 GXYKLESDKGFLEIHKTDJSDSLTYCTVKNAGSVSSCKLITIKAKTDEAKOVSTQ 26458
OY 29 ----- 28
Db 26459 KISEITPORKAVQEEISOKALNSEIIRKSEAKSQEKLAKFEASVILISEEVKSAATS 26518
OY 29 ----- 28
Db 26519 LEKSIYHEITTSQASEEVRTAEIKAFSTQMSINEGRLVKANINAGADVKWLVNGV 26578
OY 29 ----- 28
Db 26579 ELINSEYRAGVSGDQTLTIQASHRDGILTCISKTEGIVKQYDLITLSKELSDAPA 26638
OY 29 ----- 28
Db 26639 FLSQPRSONINGQVNLFTCEISCEPSPLEIWKNNLPISISNVISRSRRVYSLEIRN 26698
OY 29 ----- 28
Db 26699 ASVSDSGKTYIAKNFRQCCSATASIMVPLVEEPSREYVLTSGDTSLQSFSSQVOM 26758
OY 29 ----- 28
Db 26759 SASQEASFSSFSSSASASMTWKFPASMSAQMSMSQSFVEMSSSPMGISNMTQLESS 26818
OY 29 ----- 29
Db 26819 TSKMKACIRGIPRYEALPDSIDEGVNLVACAFTEPTEVTWMSGCRKIHSGBOG 26878
OY 30 RE 31
Db 26879 RE 26880

```

RESULT 12

Q8WZB3 PRELIMINARY; PRT: 26926 AA.

AC Q8WZB3;

DT 01-MAR-2002 (TRENDEL. 20, Created)

DT 01-MAR-2002 (TRENDEL. 20, Last sequence update)

DT 01-JUN-2002 (TRENDEL. 21, Last annotation update)

DE N2B-titin isoform.

GN TTN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20309627; PubMed=10850961;

RA Freiburg A., Trimbitas K., Hell W., Cazorla O., Fougereuse F., Centner T., Kolmerer B., Wilt C., Beckmann J.S., Gregorio C.C., Granzler H., Labelt S.;

RT "Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity.";

RL Circ. Res. 86:1114-1121(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21573839; PubMed=11717165;

RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M., McNabb M., Wilt C.C., Labelt D., Gregorio C.C., Granzler H., Labelt S.;

RT "The complete gene sequence of titin, expression of an unusual -700 kDa titin isoform and its interaction with obscurin identify a novel z-line to I-band linking system.";

RL Circ. Res. 89:1065-1072(2001).

DR EMBL; AJ277892; CAD12455.1;

DR InterPro; IPR000282; Cytok_receptor_2.

DR InterPro; IPR000719; Euk_kinase.

DR InterPro; IPR005777; Fcgt_kin.


```

OY 6 ----- 5
Db 4199 SDEKODESLRSEKESSESGTEVATVIOBEGGFIKEDGPMHITPLVTVSEEGD 4258
OY 6 ----- 5
Db 4259 IVHITITSITAKKAVNWFENKLVPSDEKFCLODQNTYTLVIDKVNTEHDQGEVCEBALN 4318
OY 6 ----- 5
Db 4319 DSGKTASAKLVVRAAPVTKRIEPLVVALGHLAKFTEIOSAPVNRQWPKAGREIY 4378
OY 6 ----- 5
Db 4379 ESDKCSIRSSKYSISLEILRTQVVDGGEYCKASNEYSVCTATLTVVPGGEKVRKL 4438
OY 6 ----- 5
Db 4439 LPERKPEKPEVYIAKSVLRKRPEEBEPKPKLEKVKKPAVPEPPPKPEVEVPTVT 4498
OY 6 ----- 5
Db 4499 KRREKRIEPTKVEIKPAIPAPAPKPEAKVTKIKPPVDEPTPIAPVTVVVGK 4558
OY 6 ----- 5
Db 4559 KAEKAPKEEAAKPKPGIKGVPKTPSPLEAERKRLPGSGGKPPDEAFVYOLKAVPL 4618
OY 6 ----- 5
Db 4619 KFKVEIKDILITSESEVGSATIECLVSPSTAITTMKDGSNITRESPKHRTIADGKRL 4678
OY 6 ----- 5
Db 4679 HIIDVOLSDAGEYTCVLRNKEKSTAKLVHEELPVRFVTKLEEVTVYKQGPLTSCB 4738
OY 6 ----- 5
Db 4739 LKNEROVVWKKDKTIYERKGRIVPGVIGLMRLTINDADDTAGTYTVENANMLECS 4798
OY 6 ----- 5
Db 4799 SCVKEVEVINDVLPPIRQGVKPKGTATFACDIANDTNIKWFKGYDELPAPENDKTEI 4858
OY 6 ----- 5
Db 4859 LRDSNHLVYKIKHAMPEDIAEYAVEIEGRKVPKALTIGREVELKPIEDVTIYEKESAS 4918
OY 6 ----- 5
Db 4919 FDAEISEADIPGQWKLKGLLRSPICEIKABGKRFILTRVKLDOAGEVLYOLMLAIT 4978
OY 6 ----- 5
Db 4979 TAILTVKEIELDEAVPLKDVYPERQARPECVLTHREANVMSKGPDIKSSDKPDITAD 5038
OY 6 ----- 5
Db 5039 GKXHLIVINDSOPDDEGVYTAIEGKKTSAFLVYIGIRLKFMSPLEDQTVKEGEMATEFVC 5098
OY 6 ----- 5
Db 5099 ELSEKGMVWVRKNDAKLTSTRTVLISSEKTHKLEMKFVTLDDISQIAOVKEIESTAQ 5158
OY 6 ----- 5
Db 5159 LKVLLEADPYFTVKLHDKTAVERDEITLKCCEVSKDVPVKWFRDGEIIVSPKYSINADGIR 5218
OY 6 ----- 5
Db 5219 KLIKIKKADLKDKGEVYCDGCTDKTKANVTBARLIKVEKPLVGEVFGETAHEIEILS 5278

```

```

OY 6 ----- 5
Db 5279 EPDVHGWKIKGQPLIASPDCELIIEBKHLILHNCQAGTGEVSPQAAKSAANLKV 5338
OY 6 ----- 5
Db 5339 KELPLIFITPLSDVKVEERDEANFECVESEBPTFWLNGTOEIVGDDRFELINKGTSHS 5398
OY 6 ----- 5
Db 5399 MYKSAFEDKAKYMFADKHTSGKLIEGIRLFLPLKDVTAKEKESAVTVELSRD 5458
OY 6 ----- 5
Db 5459 NIKVWFKNDQRLHTRSVMQDBKTHSITFKDLSIDTQIRVEAMKSEANLTVLE 5518
OY 6 ----- 5
Db 5519 GDPYFTGKLQDYTGVEKDEVILQCEISKADAFVWFKDKKEIKPSKNAVLIKADGKKRLI 5578
OY 6 ----- 5
Db 5579 LKKALKSDIGYTCDCGIDKTSGLDIEDREIKLVPLHSVEMETETARETEISEDOI 5638
OY 6 ----- 5
Db 5639 HANWKLKGBALLQTPDCEIKEGKIHSLVHNCRLDQDQGVDFQAAKSSAHLRVKRV 5698
OY 6 ----- 5
Db 5699 IGLRPLADVTVAGETAFIDCELSYEDIPVEMYLKGLKLEPSDKVVRSEKGVHTLFLR 5758
OY 6 ----- 5
Db 5759 DYKLEDAGEVOLTAQDKFTHANLTFYKEPPVEFTKPLEDQVIEGATAVLICEVSENAKV 5818
OY 6 ----- 5
Db 5819 KFKNGTEILSKKYEIYADGRKVLVHDCPTBEDIKTYTDANDKFKSCHLVNPPHVE 5878
OY 6 ----- 5
Db 5879 FLRPLTDIQVREKEMAREFECESRENAFVWFKDGAIRKGYDIISKAVRILVINKC 5938
OY 6 ----- 5
Db 5939 LIDDEAEVSCFVRTARTSGMLTVLEEAIVFTKNLANIEVSETDIKLVCEVSKPGEVIV 5998
OY 6 ----- 5
Db 5999 YKDEEIIETGRYEILTEGRRRIIVIQNAHLEDAGNYNCRLPSSRTDGKAVHLEAIFI 6058
OY 6 ----- 5
Db 6059 SKPQNLLEIBEKAEFCVCSIKESPVQMKRDKTLESQKDYVIADGKKRVLVNDATL 6118
OY 6 ----- 5
Db 6119 QDMGTYYVWGAARAANLTVIEIKRIYVPLKOTRYEEOEYVNCCEVTEGAKAKFRN 6178
OY 6 ----- 5
Db 6179 EEAIFDSKYYILQKDLVTVLRIDAHLDQANVNSLTNNGENVASAANLIVEBRDLR 6238
OY 6 ----- 5
Db 6239 IVEPLAOLEIMEKKSVYTWCKVNRNLNTLAKWTKNGEEVPPDNRSYVADVKKHMLTIKDC 6298
OY 6 ----- 5
Db 6299 GPPDEGEITVAGODKSAVELLIEAPLEFEVHELDQTYTEFDQAVFSCQLSREKANVKV 6358
OY 6 ----- 5

```

Db 6359 YRNGREIKKGGKXKFEKDSIHRLLIRDCRLDOCEYACGVEDRKSRARLFVEIPEVLI 6418
QY 6 ----- 5
Db 6419 RPPQDILAEAGADVFLAEINKEKDEYOMLRNNMVVQGDQHMSEKGIHRLQICDINP 6478
QY 6 ----- 5
Db 6479 RQGGYEFIAKDEAKALELAAPKIKIADQDLVYDVGKPLTMYVPYDAIPKAEAEWK 6538
QY 6 ----- 5
Db 6539 ENERLSTKITDTTAEQTSFRLBEAKKGDGRKIVLQNKHGAEGFINLKVIDVPGPVN 6598
QY 6 ----- 5
Db 6599 LEVTEFDEGVSILAMEPLDGGSKLIGYVERNDIKRKTWLTATDNRESCEFTVGLQK 6658
QY 6 ----- 5
Db 6659 GGVEYELRVASARNVGTGEPEYEDNPEARSKIDVPGPLMTITDYNKRGVSLTWEPE 6718
QY 6 ----- 5
Db 6719 YDGAETINVIYIELDKTISIRMDTAMTVRAEDLSAVTLDVEGOEVSFRVAQNRIGVGR 6778
QY 6 ----- 5
Db 6779 PSAATPVKVAADPIERPSPVNLTSDDQTSVOLKWEPLKDGSPILQYILIERCEBGK 6838
QY 6 ----- 5
Db 6839 DNMTIRCNKLVBELTYKGTLEKGNKYLIVYSAENAGVSDPSILGLFADDAFVEPTM 6898
QY 6 ----- 5
Db 6899 DLASAFKQGLEVIYENPITILVPSGYEPRPTATWQCFGDKVLETGRVKKKTLISAVAEIVIS 6958
QY 6 ----- 5
Db 6959 PSESDGITYTLKLENNKVTISGEIDVNVJARPSAPKELFGDITKDSVHLTWEPRDDG 7018
QY 6 ----- 5
Db 7019 GSPITGYVEKREYSRKWTYKMDVTDLEFVPLDVQGEYLFKVCARRKCGEPBAYV 7078
QY 6 ----- 5
Db 7079 DEPVNMSTPATVDPPEHYKWRDRTANSIFLTWDPKRNKGGSRIGYIVERCPGSDKW 7138
QY 6 ----- 5
Db 7139 ACGEVATKMEVTGLEGGKATYAKALNRQAGSKSPRTEIQAVDQGEAPPIFLDVK 7198
QY 6 ----- 5
Db 7199 LLAGLTVAGTKIELPATVTKPEPKITWTAKADILKQDKRITIEYENPKSTVITVDSKR 7258
QY 6 ----- 5
Db 7259 SDTGTIIEAVNVCGRATVAVENVLDKPGPPAFLITDVNESCLLTWNPBRDGGSKI 7318
QY 6 ----- 5
Db 7319 TNYVERATDSEVWHKLSYVKDINFKATILIPNKEIIFVAAENMTGVGEPEVOASPT 7378
QY 6 ----- 5
Db 7379 AKYGPDPGPPTLEPBDITKDAVTLFWCEPDDGSPITGYWERLDPDITKVRCKMK 7438
QY 6 ----- 5

Db 7439 PVKDTYHVKGLTNKKKXRFVLAENLAGPGRPSKSTEPILINDIDPFWPBGKPTVKDV 7498
QY 6 ----- 5
Db 7499 GTSYRLMTWTPEHDGAKIESYIEMLKTGTDEWVRVAEVPITTOHLLPGLMEGQESF 7558
QY 6 ----- 5
Db 7559 RYRAVNAKGESESESPDVLQREKLPPSPRMLEVINITKNTADLKNTVEKDGSP 7618
QY 6 ----- 5
Db 7619 TNYIYKROVRKRGQVDTYVKTCTYPTLTGSLYFRVAENALGSDYEIEDSV 7678
QY 6 ----- 5
Db 7679 LAKDTFTPGPYALAVDVYIKRHVDLMEPPKNDGSRPIQRYIEKKERLGTNNYAKG 7738
QY 6 ----- 5
Db 7739 TAGPCNFRVTDVIGSTEVOFVRAENAGVSHPESEPTIELIEDPTSPSPPLDLHTVD 7798
QY 6 ----- 5
Db 7799 AGRKHIALAKPPEKNGSPITGYHVEKCPGTEKAMRNSBPDKLKFVYEGVYVDKE 7858
QY 6 ----- 5
Db 7859 YVLRVAVNAIGVSESEISENVAKDPCKPTIDLEHTDIIYIGEKLSIPVEPRAVPV 7918
QY 6 ----- 5
Db 7919 PTVSWKDGKVEKASDRLTMKNDHISAHLEPKSVRADGIYITLENKLGSARASINVK 7978
QY 6 ----- 5
Db 7979 VIGLPGCKOIASDITKSKLTWEPFPGGTILHLYLERREGRRTYIPVSGENK 8038
QY 6 ----- 5
Db 8039 LSWTVADLIPNGEYFPRVAVNRKVGGEYIELEKNVIAQDPKQPPDPVDEVHNPTAEV 8098
QY 6 ----- 5
Db 8099 MTTWNPPLYDGSKIMGYIIEKIAGERMRKCNHLPILITYAKLESKEYOFRVR 8158
QY 6 ----- 5
Db 8159 AENAGISEPSRATPTKAVDPIADPKVILKTSLEVRKQDEIADASTSGSPYITWIK 8218
QY 6 ----- 5
Db 8219 DENVIYEEIKKAAPLVRRKGEVOEEPEVLPITQRLSINDSKGESQLVRDSLAPD 8278
QY 6 ----- 5
Db 8279 HGLYMKIVENDGIAFAPCTVSYLDPGPPINFEVEDINKTSVLQKWEPLDGGSEIIN 8338
QY 6 ----- 5
Db 8339 YLLEKKDKTPDSEMTIVYSLRHCKYSYTKLIEGVEYLFVRARENREGPPGVSKPLV 8398
QY 6 ----- 5
Db 8399 AKDPGPDAPOKPIVEDVTSNMLYKNNPKDNGSPILGYWLEKREYNSTHMSRVKNSL 8458
QY 6 ----- 5
Db 8459 LNLKANVDGLLEGLTYFRVCAENAGPKTSPPSDPKTAADPISPPGPIIPRYTDTSS 8518
QY 6 ----- 5
Db 8519 TTIELMEPPAFNGGGEIYGFVDKQLVGTNEMSRTKEMINVRQYTVKEIREGADYKLR 8578


```

QY 6 ----- 5
Db 8579 VSAVNAEGEPGETOPVTAEPQEPAPVELDVSKGSIQIMACKTLRIPAVTGRPVPL 8638
QY 6 ----- 5
Db 8639 KVMTEBGEJLDKRVYIDNCTKSELIINDALRKDGRVYITATNCGSKFAARVEVPD 8698
QY 6 ----- 5
Db 8699 VPGVLDKPVYTKRKKCLLNKSDPEDDGSSEITGFIIEKDKAMHTWQPIETERSKCD 8758
QY 6 ----- 5
Db 8759 ITGLLEGOEYKFRVIANKKGCGPVEIGPLAVDPLGPPSPERLTYTERKSTITIDY 8818
QY 6 ----- 5
Db 8819 KEPENSGSPIQYIIEKRKHDKPDERVYKRLCPYTSSEIYENIDEHQMTIEFVKAVNET 8878
QY 6 ----- 5
Db 8879 GSESPSIPLVAVIODEVPPLIKRLSVRGDTIKVKGEPVHIPADVTGLPMPKIEKSN 8938
QY 6 ----- 5
Db 8939 ETVIEKPTDAIQITKEVSSSEAKTELSTPKAVREDKGYTTVATSNRLGSVFNNVHEVY 8998
QY 6 ----- 5
Db 8999 DRSPRNLAVIDIKASCTLTMDAPIDNGSEITHYVIDKRDASRKKAEMEYVTNAYE 9058
QY 6 ----- 5
Db 9059 KRYGIMLIPNGQYEFRAVNRKYGISDECKSDKVYIDPVRLLPGPKPKVLARTKSM 9118
QY 6 ----- 5
Db 9119 LNSWTPIDNGSGSPITGYMLERREBSSPYKSRAPITKGLKGEFNNPRLLEGVKKY 9178
QY 6 ----- 5
Db 9179 FRAMAINAGIGPSESDPEVAGDPIPPGPSCEVDKTKSSISLGKWKPPAKDGGSP 9238
QY 6 ----- 5
Db 9239 IKGIYVMOEGTTDKRVNEPDKLITTCQCVVPNIKELKRYFRVKAVERNAGESEPSDT 9298
QY 6 ----- 5
Db 9299 TGEIPATDQEEPEVFIIDGACDCLVCKAKSQIRIPAVIKGRPTPKSSWEDGKAKKANK 9358
QY 6 ----- 5
Db 9359 DGVDHIDEDAOLETAENSVIILPECKRSHTKYSITAKNKAGOKTANCVRKVDVPGPP 9418
QY 6 ----- 5
Db 9419 KDLKVSIDITRGSGRLSKMKPDDGDRIGYIEKFTIDGKANTKVPDGGSTTVVDDL 9478
QY 6 ----- 5
Db 9479 LSEQOYFVRVARENREGISGPVETIQRTARBPYPPDPPIKLKIGLITKNTVHLSMKRP 9538
QY 6 ----- 5
Db 9539 KNDGSPVTHYIVECLAMPDTGTKKEMRQCNKRDVEELQFTVEDLVESGEYFVKAVN 9598
QY 6 ----- 5
Db 9599 AAGVSKSATVGPCCQGRDMPPSIDLKEFMEVEBGTNNVIAKIKGVFPPLTWFKAPP 9658

```

```

QY 6 ----- 5
Db 9659 KRPDNKEPVLIDHVNKKLVVDCTILVIPQSRSDTGLTTITAVNNLTASKEMRLNVIG 9718
QY 6 ----- 5
Db 9719 RQGPVPIKFESVSADQWTLSEWPPKDDCGSKITNYIEKREANRKTWVHSSPEKECT 9778
QY 6 ----- 5
Db 9779 YTIPLLEGHEVYFIRMAONKYGIGEPIDSEBETANLFSVGAPDKPTVSSVTNSMTV 9838
QY 6 ----- 5
Db 9839 NMEPEYDGSPTGYWLEMKDITSKRMKRVNRDPIKANTLGVSYKVTGLGSDYQFRV 9898
QY 6 ----- 5
Db 9899 YAINAGVGPASLPSPDAPARBDPIAPGPPFPKYTDWTKSSADLEWSPPLKDGSKVTGY 9958
QY 6 ----- 5
Db 9959 IYBKEGKEEKEKDKVEVROTKLVYTGLEGABYKTRVSAVNTAGIGEPEVTDIEM 10018
QY 6 ----- 5
Db 10019 KDRVSPDLQLDASVNRDRIYVIAAGVIRILAVSGKPPPTVWNNKERTLPQEAETTA 10078
QY 6 ----- 5
Db 10079 ISSMVIKNCQSHOGVSYSLAKNEGERKKTIVDVIDVBPVGTPLAHLNLENSCKL 10138
QY 6 ----- 5
Db 10139 TWFSPEDDGSPITMYIEKRESDRANTPYTYTTRONATYQGLQSKATFPRIAMENS 10198
QY 6 ----- 5
Db 10199 ISMGPFVETSEALVIREPTVPERPEDLEVAEVTKNVTLTMNPKYDGSSEIINYLES 10258
QY 6 ----- 5
Db 10259 RLIGTEKFKAVTNDILSKTIVKGLKEGDYEVYKRSVAVNIVGCKPSCFCTKPTCDEL 10318
QY 6 ----- 5
Db 10319 APPTLHIDFRLKLTIRVGEAFALTRYSKPKPKVSMKDEADVLEDRTIKITTPATLA 10378
QY 6 ----- 5
Db 10379 LEKIRAKRSDSGKTCVVENSTGSRKPCQVNVVDRPGPVGVPVSFDEVTKDYNVISMKP 10438
QY 6 ----- 5
Db 10439 PLDDGSKITNYIEKKEVGDKWMPVTSASAKTTCVSKLLEGKDYIFRIHAENLYGIS 10498
QY 6 ----- 5
Db 10499 DPLVDSMKARDFRVPDAPQPIVETVTKDSALVTNNKPHDGKPTNYILKEKREIMSK 10558
QY 6 ----- 5
Db 10559 RMARVTKDPIHPYTKFRVNDLLEGQYEFVSAENEIGIDPSBPSKPVFAKDIPANPSP 10618
QY 6 ----- 5
Db 10619 PYNPEALDTNCSVDLTWQPPRHDSKILIGYIEYOKVGBEMBRANRHTPESCPETRYK 10678
QY 6 ----- 5
Db 10679 VTGLEDDQTYFRVILAVNAGESDPAHVPEPVLVKDRLEPPELLIDANMAEQHIKVGD 10738
QY 6 ----- 5

```

Db 10739 LRLSAIKGVPPKVTWKKKEDRDAPIKARIDVTPVSGSKLEIRNAHEDGSIYSLEVENPA 10798
QY 6 ----- 5
Db 10799 GSKTSVAVLVDKPGPPRDLVSEIRKDSCLTWKREPLDGGSVIINTVERRDVAQAQ 10858
QY 6 ----- 5
Db 10859 WSPLSATSKSKSHPAKNAIANGNOYLFRVAENQYGRGPVETPKPIKALDPLHPPGPKD 10918
QY 6 ----- 5
Db 10919 LNHVDYDTEVSLWANKPRDGGSPITGYLVEYQEBGTQDMIKPKVTNIECVVTLQOG 10978
QY 6 ----- 5
Db 10979 KYRFRVKAENIVGLPDTTIPLECQEKLVPPSVELDKLIEGLVYKAGITVRFPAIIR 11038
QY 6 ----- 5
Db 11039 GVPVPTAKWTDSSEIKTDEHYIWEIENFSSVLTKNCLARDYGEQITVSNAGSKTVA 11098
QY 6 ----- 5
Db 11099 VHLTVLDPGPPPTGPIINILDVTPHEMTISWQPRDGGSPVINYIEKQDTRKDTMGVVS 11158
QY 6 ----- 5
Db 11159 SGSSKTKIKIPNLOKGEYFRVRAENKIGVPPLDSTPTVAAKNTSPSPGKPVVTDI 11218
QY 6 ----- 5
Db 11219 TENAAIVSWTLPKSDGSPITGYMERREVTGKWRVYNTPIADLKFRVYGLIEGNTYEF 11278
QY 6 ----- 5
Db 11279 KVFENLAGLSKSPSSDPPIKACRPKPPGPIINPKLDRSREYADLVWTKPLSDGSSPT 11338
QY 6 ----- 5
Db 11339 LGYVVECKPGTAQMNRIKDELIRQCAFVYRGLIEGNEYFRKANKANTYEGEPRDLAE 11398
QY 6 ----- 5
Db 11399 SVIAKDLHPPEVELDYTCRDVITVRVGGQIRILARVGRPEPDIWTKEGKVLVREKRV 11458
QY 6 ----- 5
Db 11459 DLIDLPVELQIKENAVRADHGKTIISAKNSGHAQSAIVNLDPRGPCONLKVNTVK 11518
QY 6 ----- 5
Db 11519 ENCTISWENPLDNGSEIINIEYERKPNCKMSIVASDVTKRLIKANLLANNEYFRVC 11578
QY 6 ----- 5
Db 11579 AENKVGVPPIETKPIIAINPIDRGEPEINAIADKGLFVYLKWRPDTGGSPNLSY 11638
QY 6 ----- 5
Db 11639 HVERRLKGSDEMERYHKSIKEIHMYADRCVENQIIEFRVOTKNEGSEDMVKIEEYVK 11698
QY 6 ----- 5
Db 11699 EDLQKPVLDLKLISGLITVKAGDTIRLEAGVGRKPPPEAATKDKDANDLHRSPPVKIDTR 11758
QY 6 ----- 5
Db 11759 ADSKFSLTAKKSDGKYYVATATAGSFAVATVNVLDKPGVRLNKIVDVSSDCIV 11818
QY 6 ----- 5

Db 11819 CWDPEDDGGEIONYIIEKCEKMAVNSIYSATVAPGTTVTRILISGNEYIFRVRAENK 11878
QY 6 ----- 5
Db 11879 IGTGPPIESKPIYAKTKDKGRPDPEVTVKSKREMTVYNNPPEYDGGKSTGYFLEKK 11938
QY 6 ----- 5
Db 11939 EKHSTRWVPVNSALPERKMKVONLLPDHEYOFRVKAENEIGIBPSLPSRPVAKDPJE 11998
QY 6 ----- 5
Db 11999 PGPPPTNFRVVDTIKHSITLGMKFPVYDGAFLIGYVEMRPKTADASPDEGMRKNAAA 12058
QY 6 ----- 5
Db 12059 QIVREFTVTSIDENOYEFRCAQNOYQIGRPAELKBAIKPKELLEPEIDLDASMKL 12118
QY 6 ----- 5
Db 12119 VIVRAGCPILRFAIVGRBAPRVTRKVGIDNVVRKGQVDLVDMAFLVIPNSTREDGK 12178
QY 6 ----- 5
Db 12179 YSLTLVNPAGEKAVYVNNRVLDTPGPVSDLKYSVYTKTSCHVSNAPPENDGSGYTHYIV 12238
QY 6 ----- 5
Db 12239 EKREADRTWSTVTPREVKKTSFHTNVLVPGNEYFRVAVNEYGVPDVKPYLASDP 12298
QY 6 ----- 5
Db 12299 LSEPPPRKLEVENTEMKNSATLAMLPLNDGGAKIDGYIISYREBQPADRWTEYSVKD 12358
QY 6 ----- 5
Db 12359 LSLVYTLGKEGKYKRRVARNVAVGSLPREAGYTEAKQDLPPKILMPDQIILKAGK 12418
QY 6 ----- 5
Db 12419 LRIEAVYGNPPTCKMKGEDEVYTSHLAVKADSSILLIKDVTNRDGSYSLAEN 12478
QY 6 ----- 5
Db 12479 SSGTDQIKRYVMDAPPPDPFDSIDADACSLSMHPILEDGSGNTNYIVKCOVS 12538
QY 6 ----- 5
Db 12539 RGDWVTAIASVTKTSCRVGKLLPQGEYIFRVRAENRFGISEPLTSPKXVAGPPFGVSEP 12598
QY 6 ----- 5
Db 12599 KNARVTKVNCQIEFVAMDRPDSGSPITGYLIERKERNSLMVANLILVNSTEYPCAG 12658
QY 6 ----- 5
Db 12659 LVEGLEYSFRITALKAGSSPPSKPTEYVTAHPVDPPGKPEVIDVYSTVSLIARPKH 12718
QY 6 ----- 5
Db 12719 DGGSKIIIGYFVACKLPGDKWYACNTAPNQIPOEERTATLEKAOYORAIAPTAVNIS 12778
QY 6 ----- 5
Db 12779 PPSPPDPVTIILAEVNPRIIDLSVAKSLLITVAKAGTNCLDAVFGKMPVTSWKKDGL 12838
QY 6 ----- 5
Db 12839 LKPAEIKAMORNICILEFSYNNKDSGDYTTAENSNGSKSAITKLVLDKPGPPASV 12898
QY 6 ----- 5
Db 12899 KINKMTSDRAMLSKEPPLDGGSEITNYIVDKRETSRPWAOVSATVPTTSCSYEKLEIG 12958

QY 6 ----- 5
Db 12959 HEYQFRIKAEKXGDPVTEPAIAKNPYDPGRCDPVLSNITKDHMTVMKPPADG 13018
QY 6 ----- 5
Db 13019 GSPITGLLEKRETOAVNMTKVNKKPIIENTLAKATGLQESTEEFRTVALNKAGPKPSD 13078
QY 6 ----- 5
Db 13079 ASKAAYARDQYPPAPAPFAYDTTRSSVSLSWKRPAYGGSPILIGLYVYKADSDMY 13138
QY 6 ----- 5
Db 13139 VRCLPQNLQKTRFETGLMEDTQYQFERYAVNKGISYDSDVDPKHYPRDILIPPEGL 13198
QY 6 ----- 5
Db 13199 DADLRKTLILAGVTMLYVPYGRPPKITWSKPNVNLDRIGLDIKSTDFTFLCEN 13258
QY 6 ----- 5
Db 13259 VNKYDAGKYILTLNESCCKEETIYVKVLDTPGPVNAVYKESKDSAYTMEPPIIDG 13318
QY 6 ----- 5
Db 13319 SPLIINVQKRDARKSKSWSTVTECSNTSFRVANLECKSYFPRVPAENEGIGDPETR 13378
QY 6 ----- 5
Db 13379 DAVKASQTPGPVVDLKVRSYKSSCSIGMKKPHSDGSRIGIVVDLFLBKNQVHKS 13438
QY 6 ----- 5
Db 13439 LSLQYSAKDLTECKEYTFRVSANENEGPSEITVAVARDVAPDLKGLPDLCLYAK 13498
QY 6 ----- 5
Db 13499 ENSNFKLPIKCKPAPSVSMKGEPLATDTRVSVSSAVNTTLIVDQOSDACKYTI 13558
QY 6 ----- 5
Db 13559 TLKNVAGTEGTLISIKVGNPGIPTGPIKFDVTAEMTLKNAPPKDGGSEITNYILEK 13618
QY 6 ----- 5
Db 13619 ROSVNNKVTGASAVOKTTFRTYRLHGMETFRSAENKYGEGIKSEPIVARRHFDV 13678
QY 6 ----- 5
Db 13679 PDAPPPNIVDVNHDVSLJTWIDPKKTGSPITGYHLBKKRNSLKNKANTPIRANDF 13738
QY 6 ----- 5
Db 13739 KYTGLTEGEYERVAIMLAGVKSPLSEBPVALDIPDPKPEVINITRNSVTLIWT 13798
QY 6 ----- 5
Db 13799 ERYDGGHKLITGYIEKRDLPKSKMKNHVNVPBCALTVDLVBGKYEFRIRAKNTAG 13858
QY 6 ----- 5
Db 13859 AISAPSESTETIICKDEYEAFTIVDPTIKDGLTIKAGDTIVLMAISILGKLPKSSMSK 13918
QY 6 ----- 5
Db 13919 AGKDIRPSDITQITSTPSSMLTIYANRKDAGEYITATNDFGKVEHVTVLADVGP 13978
QY 6 ----- 5
Db 13979 GGPVEISNVSAKATILWTPTPLEDGGSPIKSYILKRETSRLMTVSEDIOSCHHVAIK 14038

QY 6 ----- 5
Db 14039 LIOGNEYIFRVSANVNHGKEGPYQSEPVAMVDRGPPGPEKPEVSNYTKNTATVSMKRP 14098
QY 6 ----- 5
Db 14099 VDDGSEITGYHVERREKSLNVRRAIKTPVSDLCKVTGLQEGSTYERVSANRAGIG 14158
QY 6 ----- 5
Db 14159 PPSBASDSVLKMDAAYPPGPPSNPHVDTTKKSASLANGKPHYDGLITGYVESHQVIG 14218
QY 6 ----- 5
Db 14219 DEANIKDTGTALRITQFVVPDLOTKENKFRISAINDAGVEPAVIDVEIVEREMAPD 14278
QY 6 ----- 5
Db 14279 FELDDELRTLVVAGLSIRIFVPIKGRPAPEVTWKDNIMLNKRNIENTESFTLLIIP 14338
QY 6 ----- 5
Db 14339 EGNRDTGKFVNTIENPAGKSGEVNRYLDTGCVLNLRPDTIKDSVTLHMDLPLIDG 14398
QY 6 ----- 5
Db 14399 GSRITNYIEKREAUTRKSYSTATKCHKQYKVTGLSECEYFRVMAENEGIGEPLET 14458
QY 6 ----- 5
Db 14459 TEPVKASEAPSPDLSNIMDTKSTVSLAMPKPHDGSKITGYIEAQRKSDQWHTIT 14518
QY 6 ----- 5
Db 14519 TVKGLCEVYRNLTGESEYTFQVMAVNSAGRSAPRESRPVYKESQMLDELJLKGIIYQKLV 14578
QY 6 ----- 5
Db 14579 IAKAGDNKIVEIPVGRPKPYTNKGDQILKQVRNFETATSTLINIECVNSDGP 14638
QY 6 ----- 5
Db 14639 YPLARNIVGEVDVITIQVHDIPGPTGIKFEVSSDFTFSMDEPNDGVPISNYV 14698
QY 6 ----- 5
Db 14699 VEMQOTDSUTVVELATVIRITTYKATRLTGLYQFRVKAQNRKYGBCITSACTVANYP 14758
QY 6 ----- 5
Db 14759 FKVPGPPTQVATVAKDSMTISWHEPLSDGSPILAGHYERKERNGLIMQVSKALVPG 14818
QY 6 ----- 5
Db 14819 NIFKSSGLTDGIAVEFVIAENMAGKSKPSKPSBMLALDIPDPGAPVPLNIRHTVTL 14878
QY 6 ----- 5
Db 14879 KWAPREYTGFKITSYIEKRDLPNGRMKANFNINLENETVSGLEDAAEFERYIAKN 14938
QY 6 ----- 5
Db 14939 AAGAISSPESDAILTCHDVEAPKIVDVYFKDVIILKAGEARLEADVSGRPPTMEW 14998
QY 6 ----- 5
Db 14999 SKOGKELEGTAKLEIKIADFTSNLVNNDSTRDSGATTLTATNPGRAKHIFPVKVLDRP 15058
QY 6 ----- 5
Db 15059 GPPGPLAVEVTESEKCVLSNFPPLDGGAKIDHYIVQKRETSRLMTVNASVQYTKK 15118
QY 6 ----- 5

Db 15119 YKRLKGNEXIIRVMAVNNKYGVEPLESEBPVLAANPYGPPDPKNEPTVTITKDSWYCW 15178
 QY 6 ----- 5
 Db 15179 GNPDSGSEIINYIVERBDKAGQEWIKCNKKTLDJLRKVSGLTEGHEFEFRMAENNA 15238
 QY 6 ----- 5
 Db 15239 GISAPSPSPFYKACDVERKPGPPGNPKVLDJRSKSISITAMNKPIDGSGSETTGKVEIA 15298
 QY 6 ----- 5
 Db 15299 LPEDEMOIYTPPAGLKATSYITGLTENQEKIRIYAMNSEGIGEPALVPGPKAEDM 15358
 QY 6 ----- 5
 Db 15359 LPEEELDADLRKVTIRACCTLNLFPPIKGRPAPEVKARHDSLDKASIESTSSYTL 15418
 QY 6 ----- 5
 Db 15419 LIYGNVNRDSCKYLITVENSSGSAFVNVRLDTPGPPQDLKVKVTKTSVTLTWDP 15478
 QY 6 ----- 5
 Db 15479 LLDGSKIKNYIEKRESTRKAYSTVATNCHKTSKVDOLQEGCSYFRVLAENEYIGL 15538
 QY 6 ----- 5
 Db 15539 PAETAESYKASERPLPPGKITLMDVTNVSLSMEKEPHDGSRLIGYIENQTKSDXN 15598
 QY 6 ----- 5
 Db 15599 ATGATVATEATITGLQGEYSFRVSAONEKGISDRQUSVPIAKDLIIPAPKLFEN 15658
 QY 6 ----- 5
 Db 15659 TTFVLAGEDLKVDVPIGRPTPAVTMHDVNPILKQTRVNAESTENSLTITINDACREBY 15718
 QY 6 ----- 5
 Db 15719 GHVYVKLITNSGAIEITLNIYLDKPGPTGPKADEVTADSTILSNPRTKYGSGSINN 15778
 QY 6 ----- 5
 Db 15779 YIYEKROTSITTOIYSAVARTTIKACRLKTCCEYQFRIAENRYGKSTYLNSEPTVAQ 15838
 QY 6 ----- 5
 Db 15839 YPEKVPGPPTPVYTLSSRDSNEVQWNEPISDGSRYIGYHLERKERNILMTKLNKTP 15898
 QY 6 ----- 5
 Db 15899 PQYFKTTGLESGVEYEFRRSAENIYIGKPSVSECYVARDODPGREAIITVNSV 15958
 QY 6 ----- 5
 Db 15959 TLAQKKPTVDGSKITGYIYEKKELPGRMKAKASTNIIIDYHEVYGLVEDHRYEFRVIA 16018
 QY 6 ----- 5
 Db 16019 RNAAGVSESESTGAIYARDEVDPPRISMDPKYKDTIIVHAGESEFVADIDYKPIPTI 16078
 QY 6 ----- 5
 Db 16079 QWINGDDELSTAFLEIKSTDFATISLKYDAVRDYSNGYILKAKNAVAGERSVTVNKKVLD 16138
 QY 6 ----- 5
 Db 16139 RPPGPPGVVLSGYTAECTLAARPLDGGSDIINYIVERRETSRLVWTVVDANVOYJLS 16198
 QY 6 ----- 5

Db 16199 CKVTKLEGNEXTERIMAVNNKYGVEPLESEBPVAKNPVPPDAPKPAEVTITKDSMIV 16258
 QY 6 ----- 5
 Db 16259 WNERPASDGSSEILGYVLEKRDKEGIWTKCHKRLIGELRLRTYGLIENNDYERVAEN 16318
 QY 6 ----- 5
 Db 16319 AAGLSESPSAVQKACDPIKPGPPNNPKVIDITRSSVFLSMKPIYDGCCEIOGYIVE 16378
 QY 6 ----- 5
 Db 16379 KQDVSVEWTCTPPTGINKTNIIEVKLEHEYNFRICAINAGVGHADVPGLIIEE 16438
 QY 6 ----- 5
 Db 16439 KLEAPDIDLELKIINIRAGGSLRLFPPIKGRPTEVKMGKXVDSIRDAIITDVTSSF 16498
 QY 6 ----- 5
 Db 16499 TSLVDNVRDYGKYLITLSSGYSKSAFVTVHLDTPSPVNIKYTEITKDSVSIIME 16558
 QY 6 ----- 5
 Db 16559 PPLDGGSKIKNYIEKRESTRKAYSTVATNCHKTSKVDOLQEGCSYFRVLAENEYIGI 16618
 QY 6 ----- 5
 Db 16619 GLPACTADPIKABVPPGKITVDYTRNSVLSWTEPEHDGSKITQYIYEQAKHSE 16678
 QY 6 ----- 5
 Db 16679 KWSECARVKSLOAVITNLQGEYLFRRVAVNEKGRSDPSLAAPVIAKDLVTEPDYKFA 16738
 QY 6 ----- 5
 Db 16739 FSSYSVOYQDLKIEVPLISGRKPTITWTKDGLPKQTRINVTDSIDLITLSIKETHKD 16798
 QY 6 ----- 5
 Db 16799 DGGYGITVAVNVGQKTASIEIVTLKDPDPGPKFDDVSASISITSMNPPLYTGCCI 16858
 QY 6 ----- 5
 Db 16859 TNYIVOKRDTTVDVYSAVARTTLKYTKTGTEYQFRIEENRYGGSFALIESDPIV 16918
 QY 6 ----- 5
 Db 16919 AOYFKEPGPPTPPATAISNDSVLIQHEPVNNGSPVIGYHLERKERNILMTKVNKT 16978
 QY 6 ----- 5
 Db 16979 IHDQFKAONLEGIIEFRVIAENIVGVGRASKNSECYVARDODPGTPEPIWAKRN 17038
 QY 6 ----- 5
 Db 17039 ETTIGTKPVYDGSNITGYIYEKRDLPDGRMKASTNIIIDYHEVYGLVEDHRYEFRV 17098
 QY 6 ----- 5
 Db 17099 IAKNAGAIKSPDSTGPTTANDEVELPRISMDPFRDIIVVNAGETPLEADVHGKELP 17158
 QY 6 ----- 5
 Db 17159 TLEWLGRKEIESANCEIKNTDFKALLIWKDAIRIDGGYTLRASNAAGSKSPVNVKV 17218
 QY 6 ----- 5
 Db 17219 LDRPGBPBPVQVTVGTSKCSITMSPLDGGSDISHTVYAKRETSRLAWTYVASEVYT 17278
 QY 6 ----- 5
 Db 17279 NSLKYTKLEGNEXYFRIMAVNNKYGVEPLESAPVLMKNPFVLPGPXKSLVYINIAKDSM 17338

```

QY 6 ----- 5
Db 17339 TYWCNRDSDGSETIGYVERKORSIGIRWIKCKNRRLITDRLRVGTGLTEDEHEFEFRVA 17398
QY 6 ----- 5
Db 17399 ENAGVGEPPATVYKACDVEFKPGPTNAHIDYTKNSITTLAKGPIYDGGSEILGTV 17458
QY 6 ----- 5
Db 17459 VEICKADEEMQIVPTQGLNTRFEISKLTEHOEYKIRVCAINKVGLGENTSVPQVYK 17518
QY 6 ----- 5
Db 17519 EDKLEAPELDDSEIRKGIYVRAGGSARIRIPEKGRPTPEITWSREGEFTDKVQIEKV 17578
QY 6 ----- 5
Db 17579 NYTQSLDNCDBNDAGKYLKLENSGSKAFYTVKVIDTPGPONLAVZKRSALFY 17638
QY 6 ----- 5
Db 17639 WEPIIDGANKVNYIDKRESTRKAYANVSSCKSTFKVENLTGAIYFRVNAENEF 17698
QY 6 ----- 5
Db 17699 GVCYVETVDAVKAEPSPGKVTYITDVQSASLMEMKPEHDGSRVIGYVEMQPKG 17758
QY 6 ----- 5
Db 17759 TEKSIVAESKVCNAVVTGLSSQGEYQFRVAKYNEKGRSDPRVLGVPYIAKDLTIQPSLK 17818
QY 6 ----- 5
Db 17819 LPEWTYSIQAGEDIKIEPIYIGRPRNISWYKDEPLKQTRVNEBTATSTVLIKEGN 17878
QY 6 ----- 5
Db 17879 KDDCKXTVATNSAGTATENLSYIVLEKPGPVYGRDEVSADPVYISMEPPATYGGC 17938
QY 6 ----- 5
Db 17939 QISNHYVERKDTTTHVHVSATVAFITIKITLKGTETQFRIEENRYKSAPLDSKA 17998
QY 6 ----- 5
Db 17999 VTYQYPFKEPGPGTFFVTSISMDQMLVOMHEPVNDGKIKIGYHLEKESKNSILMWKLN 18058
QY 6 ----- 5
Db 18059 KTIPIQDTKFTTGLDEGLEFEKVSANIVIGKPKVSECFVARDCDPPGPREALIVIT 18118
QY 6 ----- 5
Db 18119 RNNVTLMKRPAYDGSKITGYIYEKNDLPDGRMKKASFNVLETEFYVGLVEDQRYEF 18178
QY 6 ----- 5
Db 18179 RVIAKNAGNFSEPSDSSGAILTARDEIDAPNASLDPRYKDVIVHAGETFVLEADIRGKP 18238
QY 6 ----- 5
Db 18239 IPDVVMSDKGELEETNAARKEIKSTIQKTLVYVKDCIRDDGGOYLKLSNMGSTKSIPIIT 18298
QY 6 ----- 5
Db 18299 VKVILDRPPEGPLKATGYTAEKCYLAWNPPLDDGGANISHYIIEKRETSRLSWTQVSTE 18358
QY 6 ----- 5
Db 18359 VOALNTKVTKLLPGENEYIRVNAVNTKGTGEPILESGPVACNPFKPPGPPSTPEVSATRK 18418

```

```

QY 6 ----- -HSMEXR 11
Db 18419 DSMVVTARFVDDGTEIEGYILLEKRDKEGVRTKCKNTKTLTDLRLRVGTGLTGHGHEFR 18478
QY 12 ----- 11
Db 18479 VAAENNAGVEPSEPSYFRACDALYPPGPPSNPKVTDTSRSSVSLANSKPIYDGAQYK 18538
QY 12 ----- 11
Db 18539 GYVVEKEMAADWTCTPTPTGLQKQFTYTKLENTETNFRICAINSGVGPATLPGS 18598
QY 12 ----- 11
Db 18599 VQAERIEPPEIELDADLRKVVLVASATLRLPVITIKGRPEPEVKKAKAGILTDQAOTE 18658
QY 12 ----- 11
Db 18659 VTSFTMLVIDNTRFDSGRNYLLENSGSKAFVNVHVLDSPPAVNLTIREYKDSV 18718
QY 12 ----- 11
Db 18719 TLSMEPPLIDGAKITNYIYERRETRKAYATITNCTTTRIEENLDGCSYFRVLAS 18778
QY 12 ----- 11
Db 18779 NEXIGLPAETTEPVKVSBEPLPGRVTLVDYRNTATIKMEKPESDGSKITGYVEMQ 18838
QY 12 ----- 11
Db 18839 TKSEKWSCTOVKYLEANTISGLJNGEETVFRVAANNEKGRSDPQLGVPYIARDIEKP 18898
QY 12 ----- 11
Db 18899 SYELPEHIFNVKAREQLIDVPEKGRQATVNRKOGQTLKETRVNNSSKTYVSLSK 18958
QY 12 ----- 11
Db 18959 EASKEVDGYELCVSNSAGSTVPYITVILDRPGRPGPIRIDEVSCDSITISMNPEYDG 19018
QY 12 ----- 11
Db 19019 GCQISNHYEKKETSTTHIYQAVARTSITVRLTJGSEYQFRVCAENRYKSSYES 19078
QY 12 ----- 11
Db 19079 SAVVAEYPPSPGPGTFRVHATKSTMLVTMQVYVNDGSRVIGYHLEKERSILMSK 19138
QY 12 ----- 11
Db 19139 ANKTLADTOMEVSGIDEGIMYEVYTAENIAGICKSKCHEPVAPRDCDPPGPEVTN 19198
QY 12 ----- 11
Db 19199 ITRKVSILKSNPHYDGAKITGYIVERRELFDGRMLKCNNTNOETFEVTELETDQRY 19258
QY 12 ----- 11
Db 19259 EFRVFARNADSVSESTGPIIVKDDVEPVRVMDYKFRDVIVYKAGEVLKINADTAG 19318
QY 12 ----- 11
Db 19319 RPLVYISNAKDIEIBERARTEIISTDNHTLLTVNDCIRDTGOYVILAKNVAGTRVAV 19378
QY 12 ----- 11
Db 19379 NCKVLDKPGPAGPLEINGLTAKESLSNGRQEDGADIDYIYERKETSHTANTICEG 19438
QY 12 ----- 11
Db 19439 ELQNTSCAVTKLKNNEYIFRVTVGNKYGVGBLESVAIKALDPEVPSPTSLTTSVT 19498
QY 12 ----- 11

```

```

Db 19499 KESWTLCHMRPESDGGSEISGIIIEERENSLMRVAVNKKPVYDLNFKSTGLREGEYEX 19558
QY 12 ----- 11
Db 19559 RYVAENAGLSPSESPILRAEDPVFLPSPPKPIVDGKTTITIAVVKPLFDGAPI 19618
QY 12 ----- 11
Db 19619 TGYVEKKSDPDMDKTSIOSLNGTEYTLISGLTGAEYVFRKSVKVGASDPSDSDPQ 19678
QY 12 ----- 11
Db 19679 IAKEREPEPLFDIDSEMRKTLIKAGASFTMTVPFRGRPVVJMSKPTDLRTAYVDT 19738
QY 12 ----- 11
Db 19739 TDSRTSLTIENANRMSGKYTLTQVLSAASLTVYKVLDPGPTNITVDVTRESAV 19798
QY 12 ----- 11
Db 19799 LMSDVPENDGAPYKNTHIEKRASKANVSYNNCRSLSYKVTNQBGAITYFRVSGEN 19858
QY 12 ----- 11
Db 19859 EFGVIGPAETKEGVKITEKPSPEKLGVTISKDSVSLTWLKPEDHGSRIYHYVEALE 19918
QY 12 ----- 11
Db 19919 KGORNVKCAVANSSTHNVSGLENSEYFPRVAENQAGLSDPRELLPVLKEOLEPPE 19978
QY 12 ----- 11
Db 19979 IDMKNPSSHVVYVAGSNLKVDPISGRPLPKYTLSDGVPLKATMFMETIAENLTIN 20038
QY 12 ----- 11
Db 20039 LKESVTADAGHEITANSSGQTKAFINIVLDRPGPTSPVVISDITESVTLKMEPRK 20098
QY 12 ----- 11
Db 20099 YDGSQVTLILKREISTAVTEVSATVARTMAKVKKLTGERTORITAEHRFGISDH 20158
QY 12 ----- 11
Db 20159 IDSACVTKLEPYTPPGPSPMTNNTRESITVGMHEPVSNGSAAVGYHLEMDRNSIL 20218
QY 12 ----- 11
Db 20219 WOKANKLIVIRTIIEKVTTISAGLIEFRVYAENNAVGKPSHPSEPVLAIDACEPPNR 20278
QY 12 ----- 11
Db 20279 ITDISKNSVLSMOOPAFDGSKITGYIVERDLDPGRRTKASTNTERTOFIISGLTON 20338
QY 12 ----- 11
Db 20339 SQYERFVARNAGSISNPSVVGPICTIDSYGVPVLDPLEYEVVKYRAGTSVKLAG 20398
QY 12 ----- 11
Db 20399 ISGRPAPIEYKDDKELQTNALVCVENTDILASILKADRLNSGCEYELKRNAMGSAS 20458
QY 12 ----- 11
Db 20459 ATIRVOILDKPGPGPIEFTVTAEKITTLKRPADGGAKITHIYVERETSRYVWSM 20518
QY 12 ----- 11
Db 20519 VSEHLEBCLITTTIKIKGNEYIFRVAVNKKYGISEPLSDSVAKNAFVTPGPGIDEVT 20578
QY 12 ----- 11

```

```

Db 20579 KITKSNATVWSRPIADGSDISGFLEKRDKSLGWEKVLKETIRDTROKVTGLTENSD 20638
QY 12 ----- 11
Db 20639 YQYRVCAVNAAGCPSESPSEFYKAADPIDPGPAKIRIDSTKSSITLCHMRPYDGG 20698
QY 12 ----- 11
Db 20699 SAVTGYVEIRGEEHEMTVSTGEVRTTEYVSNLPGVNYFRVSAVNCAGGEPTE 20758
QY 12 ----- 11
Db 20759 KNEPYOANDILEAPEIDVALRTSVIAKAGEDVQVLLPFGRPPPTVYTRKDEKNLSD 20818
QY 12 ----- 11
Db 20819 ANVSIENTDSSLTIPOVTRNDGTGKILLTENGVEKPSSTVSQVLDTPAACOKLOVK 20878
QY 12 ----- 11
Db 20879 HVSRTVTLMDPPLIDGSPILNIVIEKRDKRTKRWGVSHKSGSTFKLIDSEKTPF 20938
QY 12 ----- 11
Db 20939 FFRVLAENEIGICEPCEETPEYKAENVAPIRIDLSMKDSTNTSVLLSWTKRDPDGGVYT 20998
QY 12 ----- 11
Db 20999 EYVERKKGGEOTSHAGISKTEIEVQAKROSVLAEFRVAKNKGSLDPTVIGPITVK 21058
QY 12 ----- 11
Db 21059 ELITPEVLDSDIPGAOVYRIGHNVHLELPYKGRPKPSISWLKDLPEKSEFVRFSKT 21118
QY 12 ----- 11
Db 21119 EKKITLSIMAKKEHGKTYVILDNAVCRIAPVTITLGPSPKPGPIRDEIKADSVI 21178
QY 12 ----- 11
Db 21179 LMSDVPENDGGEITCYISIEKRETSQTMKMWCSAVATTFKVPMLVDAEYQFVRAEN 21238
QY 12 ----- 11
Db 21239 RYGVSOPIVSIITVAKHGRIFPGPKPIYVNTSDGSLMDADVYDGSSEVTGFNEYK 21298
QY 12 ----- 11
Db 21299 KERNSITLQKVNTPSISGHEFRATGLVEGLDQFVRYAENSAGLSPPDSKFTIANSVP 21358
QY 12 ----- 11
Db 21359 DPGTPEYIDYTRFTILKNPNPLRDGSKLVGYSIEKROGNERWRCNFTDVSBCQYTV 21418
QY 15 ----- 14
Db 21419 TGLSPGDREYFRILANNAVGTISPPSOSSGIIITDENVPPIVERGPEYFDGLIKSGES 21478
QY 15 ----- 14
Db 21479 LRKALVQGRPPRVYTFWDGVEIEKRMMEITDVLGSLFVRDATDRHGVYTVKXN 21538
QY 15 ----- 14
Db 21539 ASGSAKEIKVQDTPGKVGPIRFTNITGEKMTLMDAPLNDGCAPIYHILKRETS 21598
QY 15 ----- 14
Db 21599 KLAWALLIEDKCEASTALKLINGNEYOFVSAVNFEGVGRPLDSDPVYAQIQYVPPAP 21658
QY 15 ----- 14
Db 21659 GIPEPSNITGNSITTLTWARPESDGGSEIQYIIEERREKSTWVAVYSKRPISERTFMYT 21718

```

QY 15 -----INP----- 17
Db 21719 GLEGESEFEHWAENAGVBPASISRLKREPVNPGPPTVVKVDTSKTVLSLMS 21778
QY 18 ----- 17
Db 21779 KPVFDGMEIIGYIEMCKADJGDMHKNVAECVKTTRYVTDLQAGEEYKFRVSAINGAG 21838
QY 18 ----- 17
Db 21839 KQSCSEVGTIKAVDLTAPBLDIDANFKOTHVNRGASIRLEIAYQGRPTPTAVMSKD 21898
QY 18 ----- 17
Db 21899 SNLSRADIHTEDESFLLAVENCNRNDAGKYLTVENNSGSKSIFTVKVLDPGPPGP 21958
QY 18 ----- 17
Db 21959 TFRDVTGRSATLMDAPLDLGARLHHYVERREASRSMOYISEKTOIFRVNDLAE 22018
QY 18 ----- 17
Db 22019 VPRYFRVSAVNEKGEPEYEMPEPIVATBOPAPRRLDVVDTSKSAVLAMLKPDHDS 22078
QY 18 ----- 17
Db 22079 RITGYLEMRQSGDFWEAGHTKOLTVVERLVEKTEYEFVKAKNDAGSEPREAFSS 22138
QY 18 ----- 17
Db 22139 VAIKEPDIPTDLTGITNOLITCKAGSPFTIDVPISGRPAKVTWLEMLKETDRVS 22198
QY 18 ----- 17
Db 22199 ITTKRRTTLTWKDSMRQSGRFLLENMAGVKTESVTVVYIGRGPYGPLENSVSA 22258
QY 18 ----- 17
Db 22259 ESCVLSMGEKDGTEITNYIVERKESGTAMQVLNNSVKTKQKVTMLTKMEYSFRV 22318
QY 18 ----- 17
Db 22319 SSENREGVSKPLESADPIAEHPVPSPADTRPEVHVSAANSIMWEEFYHDGGSKIGY 22378
QY 18 ----- 17
Db 22379 WVEKKERNTILWYKKNKVCLECNKYVGLVEGLEQFRTYALMAGVSKASASRPMA 22438
QY 18 ----- 17
Db 22439 QNPVADAPGRPEVTDYRSFVSLIMSAPAYDGSKVYGIIERKPSSEVGDGRMLKCNYYI 22498
QY 18 ----- 17
Db 22499 VSDNFTVATLSBGDTYEEERYAKKNAAGVLSKSGSESTGYTCDEYAPPAKELIDARLHD 22558
QY 18 ----- 17
Db 22559 LVYTRAGSDVYLDAAVAGKREPKIINTKGDKELDCEKVSLOYTGKRAVATIKFCDRSDS 22618
QY 18 ----- 17
Db 22619 GKTYLTVKNAAGTKAVSVWVKVLDSPGCKLFSRVTOECLTAMSLPQEDGAEITHY 22678
QY 18 ----- 17
Db 22679 IVERRETSRLNWAYIVEGECPJLSYVYTRLKNNETIFRVAANKYGPVSESEPIVARN 22738
QY 18 ----- 17
Db 22739 SFTIPSPGIPBEVVGCKEHIILQMTKPSDGNESINYLVDKREKKSILMTAVNKDYV 22798

QY 18 ----- 17
Db 22799 YDTRKLTSLMEGDDYQFRTAVANAAGNSESEASNFISCREBSYTPGPSAPRVDTTK 22858
QY 18 ----- 17
Db 22859 HSIISLWTKPMYDGGTDIVGVLEKQEKODQKRVHTNATIRNTEFTVPDLKKGQKYSF 22918
QY 18 ----- 17
Db 22919 RVAANVAGNSESESHAEIPVERIEIPDELADELKQVYIRAGASLRMSVSGRP 22978
QY 18 ----- 17
Db 22979 PVITMSKQIDLASRAIIDTESYSLIVDKVNRDYDAGKYTIEAENSGKRSATVVKY 23038
QY 18 ----- 17
Db 23039 DTPGPCSVKKEVSRDSVTITWEIPTIDGAPVNNYIVERREANMAAFKVTTKCSKL 23098
QY 18 ----- 17
Db 23099 YRISGLVEGTMYFRVLPENIYIGCEPETSDAVLVSEVPLVPKLEVDVTKSTVTLAW 23158
QY 18 ----- 17
Db 23159 EKPLVDGSRRLGVVLEACKAGTERMMKVYLKPLVEHTVSLNEGRYDLFRTAQNEM 23218
QY 18 ----- 17
Db 23219 GVSEPRETVAVTVODLVLEPTIDLSMPORTINVPAGRPVELYIPIAGRPAPASNFFA 23278
QY 18 ----- 17
Db 23279 GSKLRESEVYVETHTKVAKLITRETIINDGETYLELKNYGTSTETIKYILDKGPP 23338
QY 18 ----- 17
Db 23339 TGPRIKIDIDATSLTISWPEPDLGAGLGVVEQDARHAPGMLPVSEVTSRSTFETR 23398
QY 18 ----- 17
Db 23399 LTEGNEYFRVAANRREGISYLOSLEYECNSSIRIPEPPTLQIPVSRDQMTLWYPP 23458
QY 18 ----- 17
Db 23459 EDDGGSQYTVIVERKEVRADRMVYKVPYTMTRYNSTGLTEGLEVBHRYTAINANGSG 23518
QY 18 ----- 17
Db 23519 KPSRPSKPIVAMPPIAPPKQPONPRVTDTRISVLSAMSVEDEGSKVTGYLLEMKVD 23578
QY 18 ----- 17
Db 23579 QHEHTKCNITPTKIREYTLTLHPGAEYFRYLAACNAGPQEPBAEVGTAVTMELEYPD 23638
QY 18 ----- 17
Db 23639 YELDERYOGSIFVRQGVYRLTPIKGPPEPICWTKSGODISKRAMIAJSETHTELYK 23698
QY 18 ----- 17
Db 23699 EADRGDSGYVLDVLENKCGKAVYIKRVIGSPNSBEPLEYDIOVRSVSVWRPPADD 23758
QY 18 ----- 22
Db 23759 GGADILGYTLERREVPKAAWITIDSRVGTSLVYKGLKENVEYHFRVSAENOGISPLK 23818
QY 23 ----- 22
Db 23819 SEPEYTPKTPPLNPPPEPSNPPEVLDVTKSSVLSMSRDXDGSNVTGYIERKETSTDX 23878
QY 23 ----- 22

Db 23879 WVRHKKQITTTMTVTGLVDAEYQFRIIAQNDVGLSETSPASEPVYCKDPDKPSQPG 23938
 QY 23 ----- 22
 Db 23939 EELIISIKDSVTLQMEKPECDGKEILGIVVEYRQSDSNAWKSNNKRIKDKOPTIGGL 23998
 QY 23 ----- 22
 Db 23999 LEAFEFERFAENETGLSRBRPAMSIKRLTSGEAPGIRKEMKDVTTKLGAALSCQ 24058
 QY 23 ----- 22
 Db 24059 IVGRPLDICKYRKGKELIQSRKTKMSSDGRTHPLVWTEBOEGVYTCIATNEVEVE 24118
 QY 23 ----- 22
 Db 24119 TSKRLLOATPOFNGPYLKEKYGNVSTLRLHVIYIGRPVPMATFNGOKLLQNSENT 24178
 QY 23 ----- 22
 Db 24179 TIENTENTHLMKRVQKTHAGRYKQVLSNVFGTVDAILDVEIQDKPDKPTGPIVEAL 24238
 QY 23 ----- 22
 Db 24239 LKNSAVISWKPADDGSGWITNNVVEKCEAKGAEWQVSSAISVTCRIVNLLENNGY 24298
 QY 23 ----- 22
 Db 24299 FRVSAQNTFGISDPLEVSSVYIISPEKPGAPGKPTITANTKDCVYAMKPPASDGAK 24358
 QY 23 ----- 22
 Db 24359 IRNVYLEREKERQNKWISVTTEIRETVPSYKMLEGLEVEFYKCENTLGESESEISE 24418
 QY 23 ----- 22
 Db 24419 PITERSDVPLOAPHKEBELRLNRYOSNATLYCKVTGNPKPIVKYRQGENIADGLKY 24478
 QY 23 ----- 22
 Db 24479 RIOEFGGYHQLIASVYDDATYQYRATNQGSGVSTASLEVEYPAKIHLPTELGNG 24538
 QY 23 ----- 22
 Db 24539 AVHALLRGEVYSIKIPFSGKDPVITWOKGODLIDNNGHYQYIVTRSFSLVFPNGVERKD 24598
 QY 23 ----- 25
 Db 24599 AGFYVCAKKNRFGIDQTVELADVDPPRGAKVSDVSHDSVNLWTEPASDGSKITN 24658
 QY 26 ----- 25
 Db 24659 YIVEKATTAERWLRVGOARETRVTVINLPGKTSYQFRVIAENKFGLSKSPSEPPTTK 24718
 QY 26 ----- 25
 Db 24719 EDKTRAMVDEBVEDETREVSWTKASHSTKELKXWIAADLGRGERGIHVRCVETSSK 24778
 QY 26 ----- 25
 Db 24779 TYAKFVKVGTQOVLVKEKISILNARHNLILHESFESMEELVWIFEFGISGLDIFER 24838
 QY 26 ----- 25
 Db 24839 INTSAFELNEBELVSYHQCEALQFLHSHNIGHFDIRPENIITYQTRRSSTIKIIERGQA 24898
 QY 26 ----- 25
 Db 24899 ROLKPGDNFRLLFTAPRYAPAEVQBDVYSTANDMSLGLVYVLLSGINPFLAETNOOL 24958
 QY 26 ----- 25

Db 24959 IENIMNAEYTFDEAEKREISIEAMDFVDBLLYERKSFMTASEALQHPILKOKIERVSTK 25018
 QY 26 ----- 25
 Db 25019 VITPLKRRRYHPLIKKDLNMYVSAARISCGGALNSQKGVSAKVVASIEIGVSGQIM 25078
 QY 26 ----- 25
 Db 25079 HAVGEGGHKYVCKIENYDQSTQVTVYFVGRQLENSEKYEITIEDGVALLYKDIKLD 25138
 QY 26 ----- 25
 Db 25139 DGYRCKVANDGEDSSVAELFYKGVREYDYCRRTAKKIRKRDYMLRERPEFTLP 25198
 QY 26 ----- 25
 Db 25199 LYKRTAVGENNRFQVTTVNPPEPHVTWYKSGOKIKPGDNKRYIFESDGLYQLTINSV 25258
 QY 26 ----- 25
 Db 25259 TTDDAEYTVVANKYGEDSCAKLIVLHPPTDSTLPMFRRLANECQEGSGVCFE 25318
 QY 26 ----- 25
 Db 25319 IRVSGIPPLAKNKEDQPLSLGPNIEIIEGLDYVALIIRDTLPEDNGYRWATNTAG 25378
 QY 26 ----- 25
 Db 25379 STSCQHLQVERLRTKQKQFKSKENNERHVOKIDKTLMAELISGTESVPLQVAKAL 25438
 QY 26 ----- 25
 Db 25439 REAAVLKPKAVSTYVKGEFLIEEKEKERRLPDYDPERKXKQTTIEHQRIKQFV 25498
 QY 26 ----- 25
 Db 25499 PMSDMKWKRIHDQYKPGKLDVYQKRPKRIRLSRWEOFYVMPPLRTDQYRPMKRIK 25558
 QY 26 ----- 25
 Db 25559 LSQDDLEIVRPARRRTPSPDYFYRRRRSLGDISDEBLLRPIDDYLAAMKTEBRLRL 25618
 QY 26 ----- 25
 Db 25619 EEBELGFSASPPSRPPEFELSLRYSBQAHVVEETRKDFRSTYHIPTKAEASFSY 25678
 QY 26 ----- 25
 Db 25679 AELREHQAAYRQPRQORIRIAERBEELRPVTTQHLSEYKSELDPMKREKSRKS 25738
 QY 26 ----- 28
 Db 25739 RQREVTETIIEBEYISKHQRSSSASRLRRRSLSPYIEMRPVSELIRSRPQ 25798
 QY 29 ----- 28
 Db 25799 PAEYEDOTERRSPTEPRTSPSPVSEKSLSRPERSARPDIPSRESKAKIKQKT 25858
 QY 29 ----- 28
 Db 25859 SERKYEVSQOPFTLDHAPRIILRRRSNRPVCGQNTREILANYQSKPTAEVKYHNGVELQ 25918
 QY 29 ----- 28
 Db 25919 ESSKIHVTMTSGVLLEILIDCHTDSGTYRANCINVKGEASDYSATILDVTGGDYTTYSQR 25978
 QY 29 ----- 28
 Db 25979 RDEEYRVSVPPELIRTEAVAVSSFKTSEMEASSVREKQWETREBSLSYEHSAAE 26038
 QY 29 ----- 28
 Db 26039 MKSAALEKSEKSTTRIKITTLAAILITPRSMVTEGESARESCDJDGPVPTVWTL 26098


```

Db 978 SRSRHSASASANTSYGVDIGTWICEKRRKAPSGRPVFRTRWDGQKEYYAFHOYFL 1037
QY 8 ----- 7
Db 1038 SKRSLLVLMRIIDSGEKGIAELQMLGNIGABAPNSPVIIIGTFDAVGSISPOKAEOL 1097
QY 8 -----MEKRP----- 13
Db 1098 QQLIREKFIAPDAEKIGLPVVIDSIEISCRTHNHLNIIYDTAMQLRSPSGKEPML 1157
QY 14 ----- 13
Db 1158 LQKIPASTALAEIVNVIACNLRAAGDPVLDEQYRRLYTEQNRLLNYSFRDAEELQOL 1217
QY 14 -----DINP----- 17
Db 1218 ATWCHENGVLHYDDATLRDYFLDPQWLCMDLAHVYVREIIPFAPGVNKLDDLQML 1277
QY 18 ----- 17
Db 1278 FRVVOGNGNRSGQLRIILMTYFPGFWSRLITRIADQIIEALRGVYMASODYADF 1337
QY 18 ----- 17
Db 1338 DLKTSLEDOTQNMKQGLALTYGPILFKIWEVFPQKTERTPPTDGRFKLKDGJW 1397
QY 18 ----- 17
Db 1398 SDVNLSSSILEYFPLEYVNIQEVDDNERQLAELIRPHMSQVAKLALTVDHIDLE 1457
QY 18 AWTXXKGRPY-----GRF 31
Db 1458 DWYPSLGRFVHTSEGR 1475

RESULT 14
Q96077 PRELIMINARY; PRT; 1171 AA.
ID Q96077;
AC Q96077;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Conserved hypothetical protein.
GN H1BD24.060.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid=5141.
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Align V., Hohelsel J., Brandt P., Fartmann B., Holland R.,
RA Nykatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 3 WD REPEATS (TRP-ASP DOMAINS).
EMBL: AL513466; CAD1414.1;
DR InterPro: IPR001513; Exo_endo_phos.
DR InterPro: IPR001680; WD40.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 1171 AA; 130209 MW; DQ8FE35B53232D CRC64;

Query Match 41.0%; Score 64; DB 3; Length 1171;
Best Local Similarity 3.0%; Pred. NO. 3.4e+02;
Matches 22; Conservative 1; Mismatches 8; Indels 709; Gaps 6;
QY 1 SRXHXSM----- 8

```

```

Db 397 SKTHVNSMTVDRTSNRPNEPAPPTPLISAATHRIPIRIPEVVPVTSRTANTGP11TM 456
QY 9 ----- 8
Db 457 AYPDATNTRSKPYISKGVHEIYTKYDTMDLVCGELVCTTGFTRWSLDGEIYMSLS 516
QY 9 ----- 8
Db 517 MGRGKGTAVMKPCAHVDEGKTKWIGNNGELLEADIGTQRTDPLPAHGREIIRI 576
QY 9 -----EXRP----- 13
Db 577 YRHFNQVTLDSGALNTWGPDESHTPLSNGPSFRVPRGHTESNVDELMYATGKD 636
QY 14 ----- 13
Db 637 IRYFKPTVDGKTQFQYLNALAQDAGEVIGAVIADPDWYFPGHSDGKSVYSRKYT 696
QY 14 ----- 13
Db 697 CQCIINVSQKINSILAAAGSYLMAGYNGKVAVYDTQTPMAVKREVAHENPVILAD 756
QY 14 ----- 13
Db 757 KSSSTKLDRTQYVSLGADNMRTFALLQEDMLEEEMKRRDVEYCEQNLKYLINTWAG 816
QY 14 ----- 13
Db 817 ASPHSLARYSDSDANFRLDLOTADSPDILIFQPELVLEKKAATKATKPKKESD 876
QY 14 ----- 13
Db 877 QERMSHQYDMDLIRSLIDYMGDLVHLMTSPVGLFCIFYKADLRISNLSNE 936
QY 14 ----- 13
Db 937 VKRGSLGHLGNKGAIVRFLVDTSLCFTNCHLAAGSGANGRHNDVAIIDASAFPSER 996
QY 14 -----DIN----- 16
Db 997 DPAVRIDSYVGGGDDGMLIDHELQVINDPLNVRIDTMSRDVYVAVRQGNLTKLERDQ 1056
QY 17 -----PAM-----YXRG----- 24
Db 1057 LVAKRRNGPRLRNPDELPTPEAPYKYVDGTDYTSKKRSPAWCDRLYRGGRIEQ 1116
QY 25 -----IRPV-GRF 31
Db 1117 LDYRHEVAVSDHREVSGRF 1136

RESULT 15
Q18977 PRELIMINARY; PRT; 4135 AA.
ID Q18977;
AC Q18977;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Tenascin-X.
GN TN-X.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA TISUE-SKIN;
RA MEDLINE-97426436; PubMed-9278449;
RA Elettine F., Exposito J.Y., Garrone R., Lethias C.;
RA "Characterization of flexilin, the bovine tenascin-X.";
RA J. Biol. Chem. 272:22866-22874(1997).
EMBL: Y11915; CAA72671.1;

```

DR HSP: P02671; 1FZD.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR003961; FN_III.
 DR Pfam; PF00008; EGF_14.
 DR Pfam; PF00147; fibrinogen_C; 1.
 DR Pfam; PF00041; fn3; 30.
 DR SMART; SM00181; EGF; 7.
 DR SMART; SM00001; EGF_1like; 2.
 DR SMART; SM00186; FBG; 1.
 DR SMART; SM00060; FN3; 21.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_19.
 DR PROSITE; PS01186; EGF_2; 19.
 DR PROSITE; PS00514; FIBRIN_AG_DOMAIN; 1.
 DR EGF-like domain; glycoprotein.
 KW EGF-like domain; glycoprotein.
 SQ SEQUENCE 4135 AA; 447376 MW; D903653C3444FA9 CRC64;

Query Match 41.0%; Score 64; DB 6; Length 4135;
 Best local similarity 0.48; Pred. No. 2.7e+03; Mismatches 4;
 Matches 17; Conservative 5; Indels 4048; Gaps 4;

QY 1 SRXHX-----6
 DB 56 SRLYHTYBEGGEKQVFTTHINLPSSAGCGCPKTEPPVPASHVQALRYLLELYNG 115
 QY 7 -----6
 DB 116 LKEQCTGCGCPAQAQGTQTDIRSLSHGVFLSRACSEFPGMGPTCSDEGAEGP 175
 QY 7 -----6
 DB 176 PSSPPAPSGCPDCCNDQGCVRGRCVCPGYTGPSCSMFSCPDCHGRGCVQGVCH 235
 QY 7 -----6
 DB 236 SGFSGDDCSTVRCSPRGCSQGRCEDEKVCNPGYSGEDGVRSCPRDCQNGRCENGRCY 295
 QY 7 -----6
 DB 296 CNPQYTGDDGVRSCPRGCSQGRCEDEKVCNPGYTGDDGSRITCPMDGEGRCVDGR 355
 QY 7 -----6
 DB 356 CVCHPAGIEDCSTRTCPRCRCRGCEDEKICIDPGTSGDDGVRSCPSDCNQRCEDE 415
 QY 7 -----6
 DB 416 GRVCWMPGSGPDGAGACPRDCRGRCENGVCVCHAGTSGEDGVRSCPGDCRRGRG 475
 QY 7 -----6
 DB 476 ESGRCVCMPTGTGNDGTTRACPGDCRGRCVDGRVCNPNFAGEDGSRCPGDCRG 535
 QY 7 -----6
 DB 536 RCGDVCSCVGYEGEDCGRRSCPRGCGGQCLEGRVCDDGT EGEDGVRCPRCDCN 595
 QY 7 -----6
 DB 596 RGVCQDGVCMRGFAGEDGLRVCPNCHRGRCENGRCVCDGVTGSPCAITPCPAD 655
 QY 7 -----6
 DB 656 RGRRCVQGVCHVGYSGEDCGEPPASACPGCGPRELCSAGQCVCEGREGPCAL 715
 QY 7 -----6
 DB 716 QTCGDCRGKEGECSCVQDCTAGHDCGEVPAIEGMNMLLETTVTENTRAPGV 775
 QY 7 -----6
 DB 776 DAYEIOFLPTTEGASPPFTARVPSSASAYDQKGLAPQETQVTRALRGTNMGPASKT 835

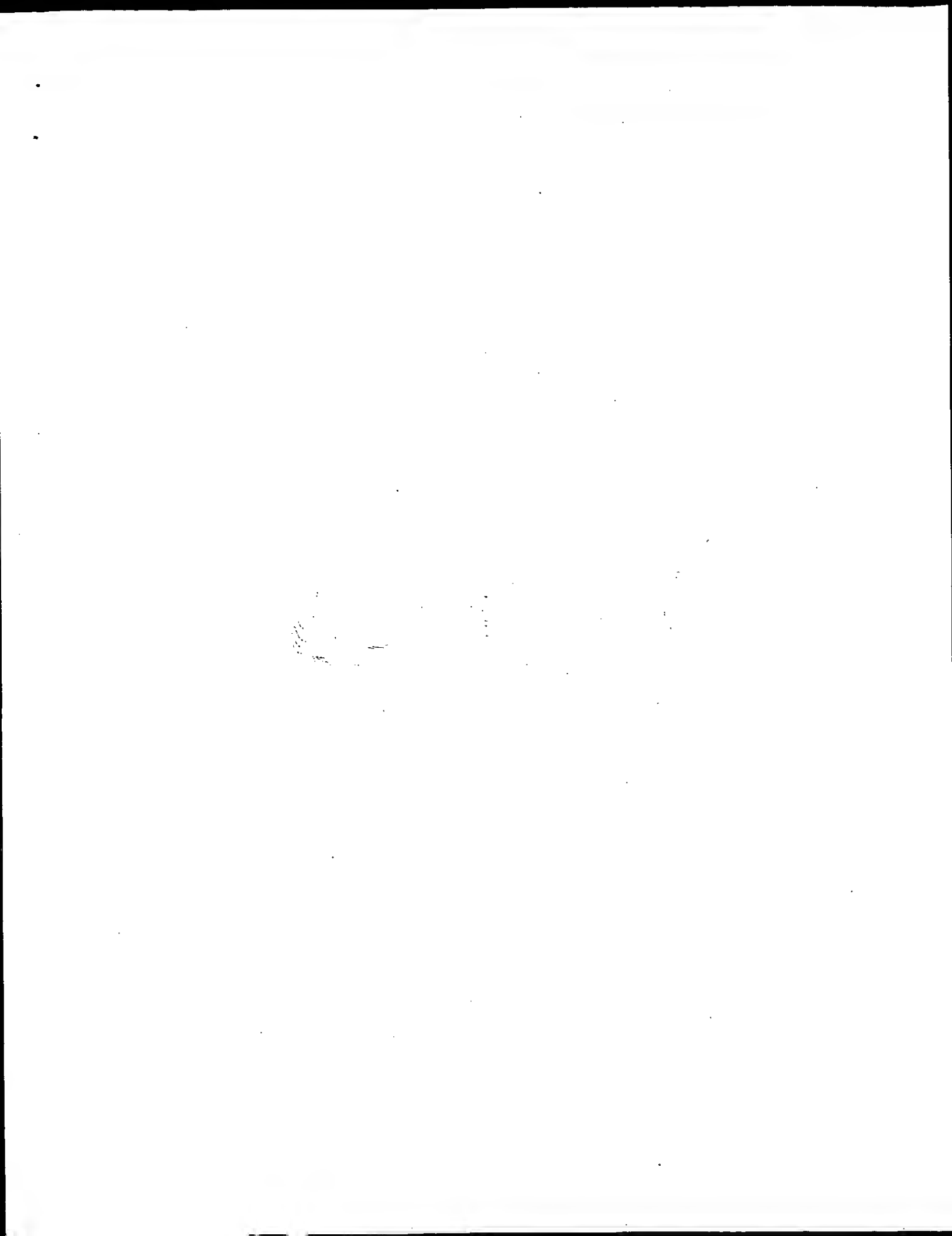
QY 7 -----6
 DB 836 TTMDCQDLRVAVPTTLELMLRQAEVDRVVSYSAGNVRLEVPSEADTILT 895
 QY 7 -----6
 DB 896 GLMPVEYVVTARGRAVSPASIRANTGSSLSGLIGATDEPPSPSTTQANAPVL 955
 QY 7 -----6
 DB 956 QQRQELAEIRYAGKDTGRLVANTAPDTFTHQLRVPEGGAHEELLPGDVROL 1015
 QY 7 -----6
 DB 1016 VSPSPESPTELISLIPPSGESPAPLIYOGIMDKCEKPGKPLAPRLGKLTVDYTS 1075
 QY 7 -----6
 DB 1076 SLLHMTVPGEFDSFVIOYKDRNPQVVPVPGPORSALISNLDYGRKRYVGLVGK 1135
 QY 7 -----6
 DB 1136 RHGPIVAEAKILSGTDPSPVTPPRLGNLWTDPTPDSLHLSWTVPBGQFQSEFVQYR 1195
 QY 7 -----6
 DB 1196 GRPQVVPVGGPDRSVLISPLDPDKRYRTLPGLANKRHGRLADGTTAPRKEPRHPE 1255
 QY 7 -----6
 DB 1256 PPERPLIGELTVAGATADSLRLSWTVAQSPDSFVQYKDAQRQAVPTGDNENALP 1315
 QY 7 -----6
 DB 1316 SLEPRKYKMLYGLHGRQVPSVAVTAPQEVIADEPSTEMEETPSTERSTKAPE 1375
 QY 7 -----6
 DB 1376 SPEKPLGELWTVGSSPDSLSTWTPQGHFDSFTVQYRGDQPRVNRVPDEGVTIS 1435
 QY 7 -----6
 DB 1436 GLEPDHRYKMLYGFHDNRQVPSVIGVTAEEETPSTEMEETPSTVEETPSMHP 1495
 QY 7 -----6
 DB 1496 STEAPEAPEPLIGELTVGSSPDSLSTWTPQGHFDSFTVQYRGDQPRVNRVGEET 1555
 QY 7 -----6
 DB 1556 EYTTIGLEPRKRYKMLYGLHGRSGSPVSTVGTAPQAEETPPATEPPKEPRIGELTV 1615
 QY 7 -----6
 DB 1616 DYTSPNSVGLMTVSRQPRSPFVQYKDRGSHVVPVADQREAVTSGLEPRKRYKMLY 1675
 QY 7 -----6
 DB 1676 GLHGQVGPLSVALTAPPPPTPEPVEPRIGELTVDTVPDVSGLSNTVAEGEPDS 1735
 QY 7 -----12
 DB 1736 FLVQYKDRDQPPVVPVATDQREVITPGLSPKRYKFLFTIGDKRSTVSVEAKTVGR 1795
 QY 13 -----12
 DB 1796 GDASGAPRLGELMTVDTPTDSLRLSWTVPBGHDSFVQYKDRGPRVSVSGHERSV 1855
 QY 13 -----12
 DB 1856 TISPLDSGRKRYFLVYGLGRRRHGRLTEGTEYRAVDAGTFRPSKRLGELQVTG 1915
 QY 13 -----12

Thu May ' 1 15:54:12 2003

seg74-plus-73.rspt

Page 57

Search completed: May 1, 2003, 14:49:24
Job time : 131 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:45:32 ; Search time 43 Seconds
(without alignments)
73.778 Million cell updates/sec

Title: SEQ74-PLUS-73
Perfect score: 156
Sequence: 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Prog. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	83	2 JC7607	prolactin-releasin
2	69	44.2	15281	2 S41309	cyclosporin synth
3	67	42.9	7463	2 T36248	CDA peptidase syn
4	66	42.3	26926	1 T38344	titin, cardiac mus
5	64	41.0	4135	2 T42629	tenascin-X - bovin
6	63	40.4	3063	2 S35505	fatty-acid synthas
7	62	39.7	1762	2 T03422	probable polyketid
8	62	39.7	2476	2 T34022	zonadhesin - pig
9	62	39.7	2611	2 T14591	actinomycin synth
10	62	39.7	2870	2 A35548	319k protein ndv
11	62	39.7	5825	2 T12117	polyprotein - fava
12	62	39.7	8363	2 T30226	polyketide synthas
13	61	39.1	1059	2 T20802	hypothetical prote
14	61	39.1	5138	2 A68695	hypothetical prote
15	61	39.1	6831	2 A88852	protein unc-22 (m
16	61	39.1	6839	2 S57242	twitichin (simil
17	61	39.1	7160	2 T27935	hypothetical prote
18	60	38.5	276	2 D70817	hypothetical prote
19	60	38.5	873	2 B41054	fasciclin II PI-11
20	60	38.5	962	2 H69157	exonuclease ABC c
21	60	38.5	1985	2 T52022	Munc13-2 - rat
22	60	38.5	2028	2 T52022	1-phosphatidylinos
23	60	38.5	2051	2 P96529	probable phosphat
24	60	38.5	3869	2 A48205	All-1 protein +GRE
25	60	38.5	3968	2 A44265	trithorax homolog
26	60	38.5	4391	2 A38096	perlecan precursor
27	60	38.5	4568	2 T08030	dynamin beta heavy
28	60	38.5	6420	2 T30283	polyketide synthas
29	60	38.5	8243	2 T31307	type I fatty acid

30	59	37.8	664	2 E83376	conserved hypothet
31	59	37.8	866	2 A95272	hypothetical prote
32	59	37.8	940	1 BVCEUA	exonuclease ABC c
33	59	37.8	940	2 H91258	excision nuclease
34	59	37.8	940	2 D86099	excision nuclease
35	59	37.8	941	2 ABL017	excision nuclease
36	59	37.8	947	2 AF0040	excision nuclease
37	59	37.8	1138	2 A48944	parasporel crystal
38	59	37.8	1555	2 S38758	amylo-alpha-1,6-gl
39	59	37.8	1779	2 T31085	xylosylase - Caldice
40	59	37.8	1807	2 T30940	vitellogenin - Pim
41	59	37.8	2273	2 T09083	hemagglutinin/hemo
42	59	37.8	3341	1 A42996	genome polyprotein
43	59	37.8	3623	2 T08618	intrinsic factor-B
44	59	37.8	4466	1 S17653	dynein beta heavy
45	59	37.8	5107	2 T29144	partial CDS - Caen

ALIGNMENTS

RESULT 1
JC7607
prolactin-releasing peptide - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C/Accession: JC7607
R/Yamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, Biochem. Biophys. Res. Commun. 281, 53-56, 2001
A/Title: Isolation and characterization of the rat prolactin-releasing peptide gene:
A/Reference number: JC7607; MUID:21092785; PMID:11178959
A/Contents: Spleen
A/Accession: JC7607
A/Molecule type: DNA
A/Residues: 1-83 (YAN)
A/Cross-references: DDBJ:AB040612; DDBJ:AB040613
C/Comments: This peptide induces arachidonic acid metabolite release from rat anterior release, and stimulation of ACTH secretion from the pituitary.
C/Genetics:
A/Genes: PRRP
A/Introns: 33/1

Query Match 94.2% Score 147; DB 2; Length 83;
Best Local Similarity 83.9% Pred. No. 1.4e-09;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 22 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 52

RESULT 2
S41309
cyclosporin synthetase - cyclosporin fungus
C/Species: Tolypocladium inflatum (cyclosporin fungus)
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Nov-2000
C/Accession: S45487; S41309
A/Weber, G.; Schoergerdorfer, K.; Schneider-scherzer, E.; Leitner, E. Curr. Genet. 26, 120-125, 1994
A/Title: The peptide synthetase catalyzing cyclosporine production in Tolypocladium n
A/Reference number: S45487; MUID:95094306; PMID:8001164
A/Accession: S45487
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-15281 (WEB)
A/Cross-references: EMBL:Z28383; NID:9440168; PIDD:CA862227.1; PID:9440169
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
A/Note: only a part of the translation is shown
C/Genetics:
A/Genetic code: SGC3
C/Superfamily: cyclosporin fungus cyclosporin synthetase; acetate-CoA lyase homology
C/Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:532-1008/Domain: acetate-CoA ligase homology <ACLI>
 F:1029-1096/Domain: acyl carrier protein homology <ACP1>
 F:1618-2069/Domain: acetate-CoA ligase homology <ACLI2>
 F:2527-2894/Domain: acyl carrier protein homology <ACP2>
 F:3115-3553/Domain: acetate-CoA ligase homology <ACLI3>
 F:4014-4081/Domain: acyl carrier protein homology <ACP3>
 F:4601-5057/Domain: acetate-CoA ligase homology <ACLI4>
 F:5506-5573/Domain: acyl carrier protein homology <ACP4>
 F:6094-6546/Domain: acetate-CoA ligase homology <ACLI5>
 F:7003-7070/Domain: acyl carrier protein homology <ACP5>
 F:7591-8042/Domain: acetate-CoA ligase homology <ACLI6>
 F:8063-8130/Domain: acyl carrier protein homology <ACP6>
 F:8652-9157/Domain: acetate-CoA ligase homology <ACLI7>
 F:9558-9625/Domain: acyl carrier protein homology <ACP7>
 F:10146-10586/Domain: acetate-CoA ligase homology <ACLI8>
 F:11055-11122/Domain: acyl carrier protein homology <ACP8>
 F:11635-12106/Domain: acetate-CoA ligase homology <ACLI9>
 F:12127-12194/Domain: acyl carrier protein homology <ACP9>
 F:12715-13159/Domain: acetate-CoA ligase homology <ACLI0>
 F:13623-13690/Domain: acyl carrier protein homology <ACP10>
 F:14213-14676/Domain: acetate-CoA ligase homology <ACLI1>
 F:14698-14765/Domain: acyl carrier protein homology <ACP11>
 F:1060,2558,4045,5537,7034,8094,9589,11086,12158,13654,14729/Binding site: phosphopantet

Query Match . 44.2%; Score 69; DB 2: Length 15281;

Best Local Similarity 0.2%; Pred. No. 8.2e+03;

Matches 20; Conservative 3; Mismatches 8; Indels 11650; Gaps 5;

QY 1 SRKH----- 4
 DB 1351 SRHOVTSFVTLAAAFRAHFRLLTGSNATIGVPSANRRPELVNIGFVNTGICRITTI 1410
 QY 5 ----- 4
 DB 1411 DENDNESLVROVRSSTTAQDNDQVPEQVYSSIMSSSRDASNPVLQMLFALHGGDD 1470
 QY 5 ----- 4
 DB 1471 LFKIOLEGEERYIPLEEVTRPDIEFHLYOGASKISGDIIFEAADLEAETIRGVSVFQE 1530
 QY 5 ----- 4
 DB 1531 VLRKGLQDQPTKMPPLIDGIPELERKGLHNYKTDYPRNKSVDVYQQQVRLSAEKTA 1590
 QY 5 ----- 4
 DB 1591 VIDSSRMSYAEIDGRSDQVAAMLROQLPAETPAVLAFSCBAVIALFGILKAGHAYL 1650
 QY 5 ----- 4
 DB 1651 PLDVNVPAAARLAILAEVKGKLVLLGAGEPSPGOSPEVSIIRIADATSPAGHSLMDG 1710
 QY 5 ----- 4
 DB 1711 KSKPTAGSLAYVITFSGTGRKGVAKIHRGVLRLVKOTNILLSLPPAOTPFMAHMSNLA 1770
 QY 5 ----- 4
 DB 1771 PDASIEWETALLNGSLVCIDRETIIDQAOLEALFRRHINIALFPALLKQCTDAAA 1830
 QY 5 ----- 4
 DB 1831 TIKSLDLVYVGDRDLTDALAKALVSEYVNAVGPENTVNSTLYSIADTEREVNGVP 1890
 QY 5 ----- 4
 DB 1891 TGRAVNSGVYMDNQNLVPLGVAGELVVTGDLARGTNPALSDRNVVYIARGQLLR 1950
 QY 5 ----- 4
 DB 1951 AVRTGDRARYRPKQGVQVEFFGRMDHQVAVRGRIELAEVEHALLSAGVHDAVVVNSQGE 2010
 QY 5 ----- 4

DB 2011 DNGVEVAFITADNETLIQDAGSSNQVWESHETITAVADITADNTIGRDTSMTS 2070
 QY 5 ----- 4
 DB 2071 MYDGTLLDKRMOEMLDNTNRTFLDQAGAHVLEIGTGQVLENIQAGLKSXYIGLEPS 2130
 QY 5 ----- 4
 DB 2131 QSAVQFNKAQPTPGLEGKAQHVAGTAMDTGRLSALSPDLIYNSVAQYFSPREVLAEV 2190
 QY 5 ----- 4
 DB 2191 VEALVRIIPGVARLIFGDMRTATYHKDELVARAVHTNGSKVTNSKVOQEVARLELEELL 2250
 QY 5 ----- 4
 DB 2251 VDPAFFSLKESLSEIEHVEILLPKNKKVNNELSSRYGAVLHIRHNNQSSHIRKINA 2310
 QY 5 ----- 4
 DB 2311 ESWIDFASQMDROGLARLKENNDASIAVFNPISKTIVERHIAKSLADHDGDTHS 2370
 QY 5 ----- 4
 DB 2371 SIIDVANTISAKERASQCPSLVDYDLVQLAEDAGFVYVSMARQSRONGALDYFFHHQF 2430
 QY 5 ----- 4
 DB 2431 TENESRALVDEPDIYKGOQARSLTNRPLQAVESRRIEAOYEOLOVLLPAYMIPARIVYL 2490
 QY 5 ----- 4
 DB 2491 QNMPLNTSGKYDRKELTLRAKYTAARTPSEELVAPRDSIAIICKERKOVLSVEVGTIDN 2550
 QY 5 ----- 4
 DB 2551 FFNVGSHLLATKILAAKLSROLNQIAVKDIFDPVYIADLAATIQDTEHNILPTSYT 2610
 QY 5 ----- 4
 DB 2611 GCVQSFQGRHLFLLDLNTGATVYLMFPAVRLRGPLVYSALAAALLAEERHETLRTTF 2670
 QY 5 ----- 4
 DB 2671 IEGQIGMOVLIHPAPKELNVIVSGEESTIQKILEKQTFPNLASERGFRLALKTG 2730
 QY 5 ----- 4
 DB 2731 EDEBILSTVNHHAISDGSVDIFQOEIGQFYSAILRGHDLAQIAPLSIOYRDPATWQRQ 2790
 QY 5 ----- 4
 DB 2791 IFQVAEHRQLAVYTKQLADNKPALLETDFKRRPMLSGRAGEIPVYVVDGLYTEKLDQFCR 2850
 QY 5 ----- 4
 DB 2851 IQVYAFYVLLAAFRAAHYRWTGHEDATIGTPIANRRPRLGILGFVNTGCMRTVDY 2910
 QY 5 ----- 4
 DB 2911 EDSEFLLVHOVRETTIAAHANODVPEQIVSNILPGSSDTSRNPVLQMLFALHSQNLK 2970
 QY 5 ----- 4
 DB 2971 VRLEGIEBEIISTAEITRDIERHLYGEARKLSIVYAADLVVPETIOSVITIFQGIQ 3030
 QY 5 ----- 4
 DB 3031 KGLCEPDMKPVASNALDGLSLSTGLLHPQDTYPCDASVQQLFKQVAVNPVDYAVRD 3090
 QY 5 ----- 4

Db 3091 ESTRLSYADLRKSDQVACLSRGIAPENFVALLAPRSCETIVALLIGVLKANLAVLPDL 3150
 QY 5 ----- 4
 Db 3151 VNPASRLAEILSEVSGMLVGAETPIPEGNAEATIRITELADATDIDHGLAAG 3210
 QY 5 ----- 4
 Db 3211 PTAASLAVIFTSGTGRKGVNVEHGIYLRKQNTITSKLPESPHMAHISMAFASV 3270
 QY 5 ----- 4
 Db 3271 WEFTTLNGSTIYCIDYFTLLSEALKEVFPQRVNVALPALLKQCLDNPALVXTL 3330
 QY 5 ----- 4
 Db 3331 SVLYIGDRIDADAAKARGLVQTAFAVNTVNTIYPIADPFIHGPIGHAVS 3390
 QY 5 ----- 4
 Db 3391 NSGAFVNDQOQITTPGANGELVTDGLARGYTSLNTRPNIYDIDEGVRAVRTGD 3450
 QY 5 ----- 9
 Db 3451 RVRVRPKDQIEFFGRINDQYKIRGRIRPAVEYVALLSHDLYTDAAVTNSQENDLEM 3510
 QY 10 ----- 9
 Db 3511 VGFVAARVADVREDESSNQVEMQTHFDSIAYADITTIQOQSLGRDPMGWSMTSMDSLK 3570
 QY 10 ----- 9
 Db 3571 KSOQOEMLDPTKSLSDSQPGHVLVGTGTGNVLENLGREGIAGSYGLEPSPATAFV 3630
 QY 10 ----- 9
 Db 3631 NKAASPPGLEDRIRVEVGTATDIDRLGDDLHGLVYVNSVAGYFSPQDYLQAVRDLK 3690
 QY 10 ----- 9
 Db 3691 VPGVERIFFGDMRSHAINRDLVAVAHVHALGDKATKAEIQREVRAESEDLLVDPAF 3750
 QY 10 ----- 9
 Db 3751 TSLTTOENTKHEVLLPKRRATINELSSRYAAVLHNDLANPAHVPSPGAWYFATKM 3810
 QY 10 ----- 9
 Db 3811 DRDALIRLNGFKISDNIAIMIPNSKTIYERTICESYVDLGDADKSDNDRVMSLSAAS 3870
 QY 10 ----- 9
 Db 3871 HAVKASLSAIDVIAQENGFEVLEISCARQWSQNGALDAVFHLPSPQSHVILDFLT 3930
 QY 10 ----- 9
 Db 3931 DMOGRBEALTNHPLNRAGSRVVERQIRRLQTLPAIMIQAQIWLXLPLNAMKVDK 3990
 QY 10 ----- 9
 Db 3991 KQLTQAOQVPAKQVSAVAPRIETELERVLCQEFSDVLDIGIMENFPLDGHSLMATAK 4050
 QY 10 ----- 9
 Db 4051 LAARISRLETHVSAKEIFDHPVCDLVLLVQGSAPHDPIVSTKYTGHPQSFAGRLM 4110
 QY 10 ----- 9
 Db 4111 FLDQAFGATRYMLAVALRGAMNHATLAALLERHGLKRTFYQNGVGMOKVAP 4170
 QY 10 ----- 9
 Db 4171 VVTEFLRIIDLSNGDDVPLTKKEQTAPFHELEPQWVALLRLGPGDYILSVVHHIT 4230

QY 10 ----- 9
 Db 4231 SDGWSVDVLFQELAGQFYSPVAVKNDPLSQOTYPLPIHYRDFALMKKPTQSEHNRQLYW 4290
 QY 10 ----- 9
 Db 4291 VEOVDASAPAEILLDPLPPSILSGQGEWSVILEGALYKNEFCRVNRYTSFVLLAAL 4350
 QY 10 ----- 9
 Db 4351 RAAYRILGSEDAITGTPIANRNPBELQIIGFEVNTCIGITVNEDEFESLVQVYST 4410
 QY 10 ----- 9
 Db 4411 ATAFARHODVPEKIVSTILPGRSDASRNPVQIMFAVHQSOKNCELKLENAHSEVYPT 4470
 QY 10 ----- 9
 Db 4471 ITTRFDLEFNLEQODKLEGSILYSTDLFEAVSQSLSVQELRLRGLHGPDPVISTLP 4530
 QY 10 ----- 9
 Db 4531 LODGIYDLQROGLDVQKTEPRDSSVYDVHHEQVSIBPDSIALIHGSEKLSYAQDRES 4590
 QY 10 ----- 9
 Db 4591 DRVAKRLNRRSPSUTLLAVLAPRSCETIIAPLOJILKANLAVLPDVKAPARIDAVSS 4650
 QY 10 ----- 9
 Db 4651 IPGNRLILGANVTPPKLOEALDFVPIRDTFTYLLDQDPTIERPSAQSLAVAMFT 4710
 QY 10 ----- 9
 Db 4711 SSGTRPKGVAMQHRNIVRYLAKNSVNAKOPPAARIAHISLAPDASSWEIYADLLGGA 4770
 QY 10 ----- 9
 Db 4771 IVCADYFTTIDPQALQETFEQENIRGAMLPSSLKQCIYQAPMISRLDILFAAGDRFSS 4830
 QY 10 ----- 9
 Db 4831 YDALQABLVSQGVNAPTEENTLSTIYVNAENDSFVNGVPIGSAVNSGAYIMDKNQ 4890
 QY 10 ----- 9
 Db 4891 QLVPAVGELVYTGDLARGYMDPKLDADRFIQTLVNGSEQVRAVYRTGDRVYRPRDFQ 4950
 QY 10 ----- 13
 Db 4951 IEFFGRMDQKIRGRIRLEPAVEBQAFNDGFVEDVAIVITPBNQREPMVAFTAKDN 5010
 QY 14 ----- 13
 Db 5011 SAREEATQIEGWEAHEGAGAYAHIEIESEALGYDPMGTSYDGTIEDKDEEMLN 5070
 QY 14 ----- 13
 Db 5071 DTKRSLDGRKAGRYLEVGTGTGMIMFNLGRSQGLERTIGLEPAPSAEFTVNNNAKSPG 5130
 QY 14 ----- 13
 Db 5131 LAGRAEVHGTAAVDVGTLOGLTSDMAVINSVAQYFPTPEVLAETIKSLVYQVGMKRIYIG 5190
 QY 14 ----- 13
 Db 5191 DKRSYANRDPFAAARAVSLADNASKDRVQKMDLEKEEBELLVDPAFALASQJQDR 5250
 QY 14 ----- 13
 Db 5251 IQHVELLPKRRKANNELSSRYAAVLHISDPLPIYKIDPEAMINFGSRLTREALQVYL 5310

QY 14 ----- 13
 Db 5311 KENENAEVAISNIPYKTYVERHIVRSIJDODANAPESMGSDWISAVTRBAQOCHTL 5370
 QY 14 ----- 13
 Db 5371 SASDILFIAEDAGFEVENSARORSONGALDAVPHILKPAEDSKVLTIKPPTDQGRPLK 5430
 QY 14 ----- 13
 Db 5431 SUTNPPLPAGSRBAELIIRBSQTLPPYMIPOITLIDMPLNANGVDRREIARAK 5490
 QY 14 ----- 13
 Db 5491 ITQSKPVEDIYPRNSYEATYCKGFTDVLGVEGITDNFNLGHSIMATKLAIRLGRQ 5550
 QY 14 ----- 13
 Db 5551 LNTRI SVNDVEDQPVVADIAAVIQNSAPHEPIKPADYTGFPQSFAGRLMFLDQLVNG 5610
 QY 14 ----- 13
 Db 5611 ATWYIMPLGIRLHSLAVDALATASALEQRHEPIRTFNEEDGVGVQVODNPKOLRI 5670
 QY 14 ----- 13
 Db 5671 IDLSTQPKDAYIAVLKHEQTLLFDLATEPGRVALLIRGESEHILSYMHIIISDGSVE 5730
 QY 14 ----- 13
 Db 5731 VLFDEMHRYSSALNQODPMEQILPLIQYNDPAAQKTEQVAHEHQOLDYTTENLADS 5790
 QY 14 ----- 13
 Db 5791 TPARELTLDPSPISGRANELPLIEGRLDRLKAFCHVQAAPPEVILLALNARARYL 5850
 QY 14 ----- 13
 Db 5851 TGAEDATIGTPIANNRBELENMIGFEVNTQMRALIEENDNFESLYNRVSTATSAFAN 5910
 QY 14 ----- 13
 Db 5911 QDVPESTYSSLLPGSRDASRPVQVILAVHSOODLKLLEGLDAVDSALSTRDY 5970
 QY 14 ----- 13
 Db 5971 FENLEHADRLSGVILAKELFKLRTIESVYSRLETLRALDQPLPLAVLPLTDGVE 6030
 QY 14 ----- 13
 Db 6031 IASKGLDVPRTDPRDANIVEVQIHVRATPDATAVNDATSIITYAQDOQSDRLATWL 6090
 QY 14 ----- 13
 Db 6091 SRRHMPETIVGLAPRSCETIIFGIMKANLAIPLIDINSPARLRSIISAVDGNLY 6150
 QY 14 ----- 13
 Db 6151 ILGSGVTAPEQENPEVEAVGQIQLAGTGLDKTQGSNARPSATSLAYITSGSTGRKG 6210
 QY 14 ----- 13
 Db 6211 VVVEHRSVTRLAKPSNVISKLPDQARVAHLNIAFASIWETATTLNGATLVCLDYHTV 6270
 QY 14 ----- 13
 Db 6271 LDORTLKEVERESITVYTLAPLALKOCVARIETIAHLDLYTGGRVGCHDAMRARS 6330
 QY 14 ----- 13
 Db 6331 VKIGMSGTPTENTVISTIEVDADENYNGVPIGKTVSNSGAVVMDRNQOLVPSGVG 6390
 QY 14 ----- 13

Db 6391 ELVVTGDLARGYTDPSLNKNRFTYITVNGESIRAVTGDVRVYRPHDQIEFFGRMDQ 6450
 QY 14 ----- 13
 Db 6451 VKIRGRIEPEVESALLSNSVQDAAVYICAPADODSGAENAVFAANTDEDETOEEE 6510
 QY 14 ----- 13
 Db 6511 ANDQVQWEINFEIAYSEVKDIRQSEVGNDEFGWTSMTDGSRLDKTKHEMLNTMAMI 6570
 QY 14 ----- 13
 Db 6571 LDAREPHVLEIGTGMWFNLAKCPGLQGYGVEPSKSAQFVNDAAQSPALKGRS 6630
 QY 14 ----- 13
 Db 6631 IYHVGATMDINKAGPIQPLVYINSVAQFPPEYTLFRVEALVQIPSVERTVFGDMRTN 6690
 QY 14 ----- 13
 Db 6691 AINRDFVASRALHTIGKANKRLVQRMIYELANEHELTPDAFTSLRTLGEKIKHYE 6750
 QY 14 ----- 13
 Db 6751 ILPRTMAATNELSKRYRAATLVHGRSREOSTIHQVSPNAIDFADGLDROTLLMLKEH 6810
 QY 14 ----- 13
 Db 6811 KDAQVAIGNIPYKTIYVERFVKSLSDEDMEQNSLDGSAMVAAYRAAOSCSLDAM 6870
 QY 14 ----- 13
 Db 6871 DVAEIAQEGYQVEVSWARQONGALDAIFHHFEPKEGARTLIEPTDYGBRVNTLT 6930
 QY 14 ----- 13
 Db 6931 NRPLNSIQSRIGTOIRKLOTLPPYMIPTSLIWDQFVNNNGKIDRELVRAIYAP 6990
 QY 14 ----- 13
 Db 6991 KPRSAATRVAPRNEIALILDEPEDVLGTEVSVLDFPDGLGHSIMATKLAARVSRRLDA 7050
 QY 14 ----- 13
 Db 7051 HISIKVFDQPVADIAASIORESAPNEPIQRPYTPAQSFAGRLMFLDQNLGATW 7110
 QY 14 ----- 13
 Db 7111 TLMPLAIRINGQLRVAALSAALFALERRNETLFTTEESDGVGVQIVGEARNSDLRVBDV 7170
 QY 14 ----- 13
 Db 7171 STGDGSEYLEVRREQVPRPDSSEPGWRVCLVKTGEEDHVLISIVMHHIIYDGSVDILR 7230
 QY 14 ----- 13
 Db 7231 GELQFYSALRGODPLIHPMLPIQYRDPAAQREAKOVEENOROLGYWSKOLYDSTPA 7290
 QY 14 ----- 13
 Db 7291 ELLTDLPRPSILSGRAGSDVDTIESSVYGALQSFCTRSVTTJFVILLJVRJIANRPLNAV 7350
 QY 14 ----- 13
 Db 7351 DDAITGFTIANNRREPLEFTVGCFTVNTQMRISIAADDNFEGILVQVRVNAVATRAAYANDV 7410
 QY 14 ----- 13
 Db 7411 PFERIVSALVPSGRNISRNLVOLNFAVQSVEDYDQVRLEGLESYVMPGEASTREDFMERH 7470
 QY 14 ----- 13

```

Db 7471 LVPBQKLTGSVLYSSDLFEQGTIONFVDIFQECRLSVLDPLTPISVLPESNAISNLES 7530
QY 14 ----- 13
Db 7531 LDLEMPYSDYRDRVTVDLFEQQAICPDGSAVWSSSQLTQAOLBQSDRAVAMTHER 7590
QY 14 ----- 13
Db 7591 HMPAESLVGVSPPSCETIITAFGIMANLAVLPDYAPDARLAALDVEGERLLLLG 7650
QY 14 ----- 13
Db 7651 AGVDPGQIPIRLSTAVIAEALSHATYDVTSIPQPSATSLAVYFTSGTGKPKGVME 7710
QY 14 ----- 13
Db 7711 HRGIVRLVROTNVNFPESSGALPVSHSNLAMDATWEIYTAVANGTVCIDRDTMD 7770
QY 14 ----- 13
Db 7771 IALNSTKREKNVRAAFTPAFLKOCIAETPELVANLEIITAGDRLDPGDANLAGTAK 7830
QY 14 ----- 13
Db 7831 GGTFNVLGHTENTAYSTFYVYVGEETPVNGVPGRGISNSHAYIIDRHOKLVPAGVAGEL 7890
QY 14 ----- 13
Db 7891 ILTGSDVARGYTDSALNKDRREYIDINGKSTWSYRTGDKARYRDRDQLEFFGRMDQVK 7950
QY 14 ----- 13
Db 7951 IRGVRIERGVELILDLHKSVLAATVYVVRPNBDEMIATITIDAEADVOTAKATIKHL 8010
QY 14 ----- 13
Db 8011 QGILPAWIPSHVILDOMPYTDNGKYDRKDLARQTYOKRSTARVPPROVEAVLC 8070
QY 14 ----- 13
Db 8071 EESYNLEVEVGTGDFEDLGSHLATAKLAARLROLNTRVSKVDFDPIILADLADI 8130
QY 14 ----- 13
Db 8131 RRGSHRDPIDATPYTGPVQSPAGSLMFLBOLNLGASWTLFPALRMGPLOTALAV 8190
QY 14 ----- 13
Db 8191 ALNALVNRHREALPTTFEDHDGVGVQIOPKSSODLRIILDSAVDITAYLAALKRQYTA 8250
QY 14 ----- 13
Db 8251 FDLTSEGRVSVLKLDDDDIISIVMHLISGWTVDVLRQELGFTSAIRGOEPLSQ 8310
QY 14 ----- 13
Db 8311 AKSLPIQYRDFAVWQENQIKBQAKQAKXWSQALDSTCEFTDLPRPSIISGEADAY 8370
QY 14 ----- 13
Db 8371 PMVLDGVYQIOLDPFCRTHQVTSFVLLAFTAHYRLTGLDAVGTPIANNRRLPELG 8430
QY 14 ----- 13
Db 8431 LIGFVVTQCKRAALSETETFPESLVOOVRLITTEAFANODVPEQIVSTILPGSRDTSRN 8490
QY 14 ----- 13
Db 8491 PLVGVFALQSODLGRIOLEGTDEALETPISTRLDLEVHLFOYGVKLSGSLISTDLE 8550
QY 14 ----- 13
Db 8551 EVETTINGIVDFLEILRGLRQPKORLAMPITDGIITKLKDOGLLVAKPAYPRESSVID 8610

```

```

QY 14 ----- 13
Db 8611 LFRQVAAAPDALAVMSSLTLYADLDGOSNKLAMRLCORNAPELTIVAVRAPSCLTI 8670
QY 14 ----- 13
Db 8671 VAEGLVKANLAVLPDVNAPARIBALISAVGKLVVQAHGPELGLTMADELYOJD 8730
QY 14 ----- 13
Db 8731 EALASSSGDHEQIHASGPTAVSLAVMTSGSTGRPKGVMTDHRSIINLVKNSDVATL 8790
QY 14 ----- 13
Db 8791 PTPVRANVSNLAFDISVOEITYTALLNGTIVCLDYLTLLDSKILYVFEVZOVNAAKT 8850
QY 14 ----- 13
Db 8851 PVILKOCIGMPALISRLSVLFWGDRDLDAHVAASGLIDAVYNAVGPTEMGOSWYK 8910
QY 14 ----- 13
Db 8911 KYDVNRPVNGVPTIGSITNSGATYMDGNOQLVSPGVMEIYVVGDLARGYTDALDED 8970
QY 14 ----- 13
Db 8971 RFVHTIIDGEENIKAYRTGDRVYRPKDFEIEFGMDQOVKIRGHRIPEVEHALLGH 9030
QY 14 ----- 13
Db 9031 DLYHDAVYLRRPANOPEMIAFTSQEDETLEQESNKQVOGWGHEHEDVSHYADIKDLD 9090
QY 14 ----- 13
Db 9091 TSTFGHDFLGWTSMDGVDPVNEKKEJLDETTASILNRPFGHILEGAGTGMILSNIG 9150
QY 14 ----- 13
Db 9151 KYVDGLQYVGLBPAPSAITFVNEAVKSLPSLAGKARVLTGALDGLDKNEIOPELYVI 9210
QY 14 ----- 13
Db 9211 NSVAQYPTSEYLIKVKAVEVPVSKRVFPEDIRSOALNRDPLARAVRALGDNRKEQ 9270
QY 14 ----- 13
Db 9271 IREKIALESESEBELVDPAFVYSLSQLPNTKHYEVLPLMKATNELSSTRYAAYLHS 9330
QY 14 ----- 17
Db 9331 HMEEOILLIDIDPAMVDPAATOKDSGLRNLQGRDQVIAVGNIPYSKTIYERIM 9390
QY 18 ----- 17
Db 9391 NSLDDHVSJLDTGTSHTSDARSAAALCTSFDAIPALTOIAKEBGFVELSWANORSQNAL 9450
QY 18 ----- 17
Db 9451 DAVFRLATDANCERSVLVHFPIDHQQLFTLINRPIQRAQSRIESOYFEALQTLAP 9510
QY 18 ----- 17
Db 9511 AYNIPSRIIVLQMPFNANGKYDRKOLARQAVVAKRAVSARVAPRNDTEIVCEEYAD 9570
QY 18 ----- 17
Db 9571 IIGTVGITDNFEDMGSHLATAKLAARLSRRLDRTVTVKEVDFKPVLLADAASIEGST 9630
QY 18 ----- 17
Db 9631 PHLPASSVYSGPVQSVTAQGRLEWFLDQFNINATVYHMSLAMBRLGLPILMNDALDVALRAL 9690

```

```

OY 18 ----- 17
Db 9691 EORHETLRTTEAKDIGOVVHEAMKRLKVLDSLKNEKEHVALENDOMRPTLASE 9750
OY 18 ----- 17
Db 9751 PGRKGHLARLAPTEYILSLVHMHNSDGMVLDLRLQELGQFYSALNGRDLPSQVKPLPI 9810
OY 18 ----- 17
Db 9811 QYRDFAMQKEAQAHERQLAWENQLADSTPGELLTDFPRPQTLGAGVLPVTIEG 9870
OY 18 ----- 17
Db 9871 PYTEKLKFSKEROVTLFVSLLTAFRATHRLTGAEDATIGTPIANNRPELEHIIIGFV 9930
OY 18 ----- 17
Db 9931 NTQCMRLLDYGFSTFESLVQHRSVAVDAYSNODIPEERIVSALLPGSRDASRPLIQLM 9990
OY 18 ----- 17
Db 9991 FALHSQDPLGNITLBLEHEHRLPTSVATREDMEFHLFOEPNKLSGSILFADDELFOPETIN 10050
OY 18 ----- 17
Db 10051 SVTVFQELARKGLDQPVASISMPVLDGLIDLEKLGLELESNFRDYSVQVDFRQV 10110
OY 18 ----- 27
Db 10111 AANPAPAVVDSSTSMYTSILQKSEQLAAVLHAQGLRPSLICVMARSPETIVSLGFI 10170
OY 28 ----- 27
Db 10171 LKAGTAVLPDVNSPANITQPLSEVBGKRLVLLGSGINDQSRDMEVETARIDITTY 10230
OY 28 ----- 27
Db 10231 KVERSDPMRSRATSIAVYIFSGSTGRPKGVMLEHRLILTVQSVTSQPLQDLRAH 10290
OY 28 ----- 27
Db 10291 ISNLAFLASLWEIETPALINGGALCIDYFTLLDSQALNTEFEKARVANTLEAPALKECL 10350
OY 28 ----- 27
Db 10351 NHAPLLEFDLAVYIGDRLATDAKIQALVKGYVNAVPTENTYASTYRLLDSEY 10410
OY 28 ----- 27
Db 10411 ANGVPIGNAVSSGAYIMDQKRLVPQMGELVSGDLGARGYNTGLNDRFVDIVIN 10470
OY 28 ----- 27
Db 10471 DOKARAVPTGDKTRYRPDGSIEFFGRDQOVKIRGARVEPAVEQAHLKALIHDAVV 10530
OY 28 ----- 27
Db 10531 VQAVDQGETEMIGFVSNASDRFSBGESEITNOVQEWEDHFSTAVAGIEALDQATLGDF 10590
OY 28 ----- 27
Db 10591 TSWTSMYNGNLTDKAEEMELDWTQSLDKEDARPCALIGTGKVLFNPKNDGLESY 10650
OY 28 ----- 27
Db 10651 VGIEPSKSAALFVDKAADPQLQGTQILVGTAEKILVADPHDDVVINSVAQVPSR 10710
OY 28 ----- 27
Db 10711 SYLVQJASBELHMTSVKTIFFGDMKSRATNDFLVSRLALYTLGDKATQDQJROVARLEE 10770
OY 28 ----- 27

```

```

Db 10771 NEDELLVDPAFETSLISQPGCAVKHVEILPKRMPTSNELSYRAVAVLHICRDEGRNRY 10830
OY 28 ----- 27
Db 10831 GRRVHSEENANTIDFASSGMDHVALVOMLDERDKATVAGINIPHSNTINERHFTSLDT 10890
OY 28 ----- 27
Db 10891 EGRGTAQDSLDSANOSATKAMAARCPCLSVTEIVEIGQAAGFVYEVMAKORQHGALD 10950
OY 28 ----- 27
Db 10951 VVFHLEDDRGVRLINPEPTDERLPSTGLSRPLQRIQNRRESQIREQLQTLPPYM 11010
OY 28 ----- 27
Db 11011 VPSRLVLEMRPLANSKYDRKELARKAKTQITIPSATRVAPRNDLEAVLCDEFQAVLG 11070
OY 28 ----- 27
Db 11071 VYGVADNFEELGSHMATKLAHLRSRLDTRVSVKDFINQPILODLADVQTSAPHE 11130
OY 28 ----- 27
Db 11131 ALPSTPYSGPVQSFQGRLEFLDQILNLSYTHMPLASRLRGPLRIBALQSLATIEAR 11190
OY 28 ----- 27
Db 11191 HESLRTFEEBQDGPVQIVRAARNQOLRIIDVSGEDAYLAALKQEBQADFDLTAEPGR 11250
OY 28 ----- 27
Db 11251 VALLRGPDHVLIVMHHIISQGSVDLRLQELQGLVSNASSQAPLPIQYRDAIYQK 11310
OY 28 ----- 27
Db 11311 QDSQIAEHOKQIMYMERQLVNSKPAELADFTPRKALSGDAVLPLEIDQYQMLRSFC 11370
OY 28 ----- 27
Db 11371 RANHTSFVALLAAPRAATYRLTGADANTIGSPINRNPBELGLICGVNTQCLRIYK 11430
OY 28 ----- 27
Db 11431 SEDTFDLVQARETATEAQDNDVPFERTVSKVASSRQTSRNPVQVMAFVSHQHDIG 11490
OY 28 ----- 27
Db 11491 NIRLEVEGRPVSMASSTNPDAMHLEFDQMLGQNVFSDKLPSEITRSVAVAFQETL 11550
OY 28 ----- 27
Db 11551 RRGLANPHANLATPLTDLPLSLRSLCLQVNPDPYPRDASVIDVFEQVASIPKSTAVID 11610
OY 28 ----- 27
Db 11611 ASSQLTYTELDERSQOLATRLRQVTVPELVGLAPRSCETIIAFLGIKANLALPLD 11670
OY 28 ----- 27
Db 11671 VNAPAGRIETILSLFGNRLILIGSDTOAVKLANSVRFTRLSDALVESGSPTEELSTR 11730
OY 28 ----- 27
Db 11731 PTAQSLATVAFISGSTGVKGVVVERHGITRLVKNSNVAKOPAPAAALHLSNIAFDASS 11790
OY 28 ----- 27
Db 11791 WEIYAPLNGTVCIDYTYTIDIKALEAVFKOHIRGAMLPALLKQCLVSAPTMISL 11850
OY 28 ----- 27

```

Db 11851 ELIFAGGRLLSSDALLARRAVSSGYNAYPEENTVLSIHNIGENAFSGNPGISNAV 11910
 QY 28 ----- 27
 Db 11911 SNSGAFVMDQNOOLVSAGVIGELVVTGDLRGYTDSKLRVDFYITLIDGNRYAVRTG 11970
 QY 28 ----- 27
 Db 11971 DRYHRRKDDQIEFFGKMDQIKIRGRIRPAVEQALARDPAISDAVITOLDDEEPE 12030
 QY 28 ----- 27
 Db 12031 LVAFSLKGNANGTVNGVSDQEKIDGQDHALIMENKIRHNLQALLPTVMIPSKRIHV 12090
 QY 28 ----- 27
 Db 12091 DQLPVNAKGIDRNELAVRAQAPRTSSVSTYVAPRNDIETICKERADILSVRGITDN 12150
 QY 28 ----- 27
 Db 12151 PFPLGGSLTATKLARLSRLDTRVSVNDVDTPVVGQLAASIQGGSTPHFAIPALSHS 12210
 QY 28 ----- 27
 Db 12211 GPVQSPAQGRMLFIDRFNLNAAMYIMPGVRLRGPLRVDAQTALRALERHLLRTYV 12270
 QY 28 ----- 27
 Db 12271 EEDGVGMOIVHSPRMKDVCVVISGAMEDLAKLKEBQAPNLSTEVAVRVALFRAGEN 12330
 QY 28 ----- 27
 Db 12331 HHILSTVMHIIIDGKSVDFIQOELAOFTSVAVRGHPDLSQVRPLPIHYRDFAVMORODK 12390
 QY 28 ----- 27
 Db 12391 QVAVMESOLQYWTIEQLADSTPAELISDFENRPEVLSERAGTVPLVIDEYVERKLSFCRHH 12450
 QY 28 ----- 27
 Db 12451 QVTSFVLLAFAVRAHYRLTGADATIGPPIANRNPBELEDLIGFVNTQCMRLALEHD 12510
 QY 28 ----- 27
 Db 12511 NPLSVYRRVSTAAASAFENODVPEFRLVSALLPSSDASRNPVLQMLFVHSGRNLGKLQ 12570
 QY 28 ----- 27
 Db 12571 LRGLEGEPFYTATRFDEVEHLEFQDQGLAGNVFAADLFEAAIIRSVVEVPHETLRGG 12630
 QY 28 ----- 27
 Db 12631 LDOPDIAISTMPLVDGIALALNSHNLPAVEDIEPDATEASVVDVQVQVAVNDALAVTD 12690
 QY 28 ----- 27
 Db 12691 TSTKLTVAELDOOSDEVAANLSTQKLPASIVYVLA PRSSEITVACIGILKANLAYLPMD 12750
 QY 28 ----- 27
 Db 12751 SNVPEARQAILSEIPEGEKVFLLGAGVPIPDNKTADVRAVFTSDIVASTDKSYSPGTRP 12810
 QY 28 ----- 27
 Db 12811 SASSLATVIFTSGSGTRPGKVMYEHKGVISLVKQNSRIPOSTLRMAVSNLAEDASVWEI 12870
 QY 28 ----- 27
 Db 12871 FTTLLNGSTLFCISTYFTVLDKRALSAASDHRNITLPLPALLKQCLADAPSVLSLEST 12930
 QY 28 ----- 27
 Db 12931 YTGDRLDGADATKVDLVKGRAYNAVGPTENSVNSTIYIEHETPANGVPIGTSLGPKS 12990

QY 28 ----- VGRF 31
 Db 12991 KATVMDQDQQLVPAGVWGLVYAGDGLANGYTDPSLNTGRF 13031

RESULT 3
 T36248
 CDA peptide synthetase I - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
 C:Accession: T36248
 R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrall, B.G.; Rajandream,
 submitted to the EMBL Data Library, March 1999
 A:Reference number: 221602
 A:Accession: T36248
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-7463 <SAU>
 A:Cross-references: EMBL:AL035640; PIDN:CAH38518.1; GSPDB:GNO0070; SCOPDB:SCOE63.03C
 A:Experimental source: strain A3(2)
 C:Genetics:

A:Gene: cdapsI, SCOE63.03C
 C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:516-1074/Domain: acetate-CoA ligase homology #status atypical <ACLI>
 F:1090-1158/Domain: acyl carrier protein homology <ACP1>
 F:1715-2184/Domain: acetate-CoA ligase homology <ACLI2>
 F:2200-2268/Domain: acyl carrier protein homology <ACP2>
 F:2804-3249/Domain: acetate-CoA ligase homology <ACP3>
 F:3265-3332/Domain: acyl carrier protein homology <ACP3>
 F:4323-4746/Domain: acetate-CoA ligase homology <ACP3>
 F:4765-4830/Domain: acyl carrier protein homology <ACP4>
 F:5363-5786/Domain: acetate-CoA ligase homology <ACP5>
 F:5802-5870/Domain: acyl carrier protein homology <ACP5>
 F:6401-6868/Domain: acetate-CoA ligase homology <ACP6>
 F:6884-6951/Domain: acyl carrier protein homology <ACP6>
 F:1122,2232,3297,4794,5834,6916/Binding site: phosphopantetheine (Ser) (covalent) #st

Query Match 42.9%; Score 67; DB 2; Length 7463;
 Best Local Similarity 0.5%; Pred. No. 3.5e-03;
 Matches 17; Conservative 2; Mismatches 8; Indels 3127; Gaps 3;

QY 4 HXSMEXKRP ----- 13
 Db 3537 HHSVYLDPTPHLRDPARDHATAGETLAEADDTTALTLTVPGVCHAVNDVFLSTF 3536
 QY 14 ----- 13
 Db 3597 ALAVAGWNRGRGEDADAPVILDESHGRHEAVNGVELSRTAGFTSKYPVRLAPAGAS 3656
 QY 14 ----- 13
 Db 3657 GDGSAIRAIKAVKEDQKIVPGDGGLRLINPRTAALALPLPEGFNYLGRIGQEG 3716
 QY 14 ----- 16
 Db 3717 IDEAPWTIEGGDVAGIDANPLABPVAVNAVAREADSTRARMTYSRTALEPDTORL 3776
 QY 17 ----- 16
 Db 3777 ADTWERLLRLVEARQAGAGLTPSDIAMPALAOELEDLEHTVPGIQLDILPLAPLQEG 3836
 QY 17 ----- 16
 Db 3837 FLFLNLVDENARDVYVGGLAFDLGSPDGTGRRAAGALLRRANLNAAGFQVATGTWQ 3896
 QY 17 ----- 16
 Db 3897 VVPALEPDWRECDLTDRADEAREDAEGRILAAGDRERRFDLTSPPLMRTALRLSADRY 3956
 QY 17 ----- 16

```

Db 3957 RLVMTNHHLLDGKSNPLMLWQELTELYVSGDVPVSLPVPVRYRDLMLGARDRDARDA 4016
QY 17 ----- 16
DD 4017 WRSLSGIDEATLLAPDAGPAEAPLIGIPGLDRDATALSANGVMTNTVOGANA 4076
QY 17 ----- 16
DD 4077 LALAQTGRDQVVGATVSGRPPELPGVESMIGLFINTELPVRLDQAEPLGDLFRLLON 4136
QY 17 ----- 16
Db 4137 EOARLIDHQWPIGLADIQHAGHELEFDTAMVFQNPYSADTTSKOLGILVAGYDAVEST 4196
QY 17 ----- 16
DD 4197 DPAVNLVAHPRDALARLRLDYL RADACAGDLYRSLADRMRLVLEALVTDSDRPAHLDOTLD 4256
QY 17 ----- 16
Db 4257 PAVRERYLVENGAPTOLPOTPLHLLSEQARLPPDAVAVCGDTSILTVAELDGANOLA 4316
QY 17 ----- 16
DD 4317 RHLGEGIAEDFVALALAKSLDAVISMLAVLKTGAAYLPIDDPYPAERITTMILDDARPA 4376
QY 17 ----- 16
Db 4377 LITLPEVPERYTGSHVYANTDERKSPGSAARAAMVITTSSTGRKXGVILEHALATY 4436
QY 17 ----- 16
Db 4437 LHRARNTYAMGVTVLHSPPLAFDLITLALMTPLSGGVHLNLSLESDQPSLIKATPS 4496
QY 17 ----- 16
DD 4497 HLPLLTLPETASPSHTLLIGGEALHTDHLVTWTHQHPGVQIINAGPTESTVNTDHHV 4556
QY 17 ----- 16
Db 4557 GEDRPDPVPIGRFANTQVYVILDSALRPVAPGVTELYLAGBQLANGYIGRPALTFERF 4616
QY 17 ----- 16
Db 4617 TANPSSPGARMTGTDLAMHMHGRLTYDGRADHQIKLNGHIEPGELEANTLAQTGI 4676
QY 17 ----- 16
Db 4677 TOATVOLREDPGDOKLVAIVLVNDSTBYDEKTVRDLATSLADPYWPSALVTILDALPLT 4736
QY 17 ----- 16
Db 4737 PNGLDRTALPAPAYSASTAGRAPRTFREYICTLFAEVLGYDLVTIDNFFDLGSHSL 4796
QY 17 ----- 16
DD 4797 ATRVSRTRALGVELSIKQLEPTVAGLAEALDASGTATALTANPRERIPLSYAOQ 4856
QY 17 ----- 16
DD 4857 RLWFLHOLEGSPSANTYVNLTLRIGCALDVADALRAISDVVARHESLNTVTEDEGAIQI 4916
QY 17 ----- 16
DD 4917 VLPVEASTFTTVDAVEEIGDLDEAVGHCFDLAQELPARISLFFVRSREHVLILLIH 4976
QY 17 ----- 16
DD 4977 HIASDANSRAPIADOLPAAYAAVARSEAPWAPLTVQADYALMOQELIGDFTDADSIAG 5036
QY 17 ----- 16
DD 5037 ROLATYWKQOLAGLPEQOLDPTDRPRPAVAGYSGDRVFPVTELETRULTELARATNTSAF 5096

```

```

QY 17 ----- 16
DD 5097 WYQAAVAVLLTRLGAGEDIPIGTPVAGRIDDADDLIGLFINTELVKNTDSGPTERRL 5156
QY 17 ----- 16
Db 5157 LDRVNDTDLAAVAHQDLPEERLVEALNPARULSHHPLFQVLLFTFNNDHGKALNDISELP 5216
QY 17 ----- 16
Db 5217 GLTVALREVORTSKPDLSPGFAESFDTSRROGIEALDESTEILLDRSAQALADRFLR 5276
QY 17 ----- 16
Db 5277 VLEAVTTAPDRIGAVELMDPAERERYLVENGAPTOLPOTPLHLLSEQARLPPDAVAV 5336
QY 17 ----- 16
Db 5337 VCGDTTLTYAEIDRRANOLARHLGEGIAEDFVALALAKSLDAVISMLAVLKTGAAYLP 5396
QY 17 ----- 16
Db 5397 IDPDYPAERITTMILDDARPALTLAPLEPPASTDSPTSEITDVERKSPGSAARAAMVIT 5456
QY 17 ----- 16
Db 5457 SSGTGRPKGVILEHALATYLRARNTYAMGVTVLHSPPLAFDLITLALMTPLAGGV 5516
QY 17 ----- 16
DD 5517 HLTSLAEAVOPSLAKAPSHLPLTLTPETASPSHTLLIGGEALHTDHLMTWTHQHPGA 5576
QY 17 ----- 16
Db 5577 QIINAGPTESTVNTDHHVSEDTPDGPVPIGRFANTQVYVILDSALRPVAPGVTELYL 5636
QY 17 ----- 16
Db 5637 AGBOLANGYIGRPALTFERFANPSSPGARMTGTDLAMHMHGRLTYDGRADHQIKL 5696
QY 17 ----- 16
DD 5697 KGHRIEPELETTLTAQNGITQITVOLREDPGDOKLVAIVLVNDSTBYDEPTLNDALAS 5756
QY 17 ----- 16
Db 5757 ALPDYKPSAYVTLDMALPLTPNGKIDRTALRPAYSASTGTGRTPRPRELLICTLFAEVL 5816
QY 17 ----- 16
Db 5817 GVDLVITIDNFFDLGSHSLATRLVSRATLGVLSKOFEPETPIAGLSGAFDRAGRA 5876
QY 17 ----- 16
Db 5877 RAALTARPRERIPLSYAOQLMTLHOLEGSPATYNIPTTLNGTLDPLAQSLANDLL 5936
QY 17 ----- 16
DD 5937 ARHESLRTTYTEDGBRPVYIHANERGMPLGVNDTGEHELDMLSAGVHNHAFDLTAGIP 5996
QY 17 ----- 16
DD 5997 VRATLRISEGEHVLILLIHITADMSRPIGHEDLAAVSARCADVAPWEPLPVQAD 6056
QY 20 ----- 19
DD 6057 YALMOREVLGDRGDADAPAGROLANTROLADLPBOLDPTDRPRPAVASODGDRVAFSL 6116
QY 20 ----- 19
Db 6117 DADLVYRLTELARATHTSTFWVQAAVAVLLTRLGAGEDIPIGTPVAGRITDANTENLVGF 6176

```

Oy 20 ----- 19
 Db 6177 EVNVLVINDTSGNPTRELLLETRRDVLAAYAHODLPFERVYALHPARTLAHHPLEQV 6236
 Oy 20 ----- 19
 Db 6237 MLITSTAPDPAISALPLGLVGAERSRGAKAKVDLAFALAEVNDGGRSTGLTGALDFR 6296
 Oy 20 ----- 19
 Db 6297 TDLFDRSTARSLVERFVTELEAVVADPGVLSRVPLTNGSRSLDRGTPLLEGIDAT 6356
 Oy 20 ----- 19
 Db 6357 LPELFAEQALTPGAPALVNGTGVSTAEALDRLTNRLARLRQGVRRPPTVYVLMERSP 6416
 Oy 20 ----- 19
 Db 6417 AHVAVTALAKAGAYPDLHDTPYLDNRHNVADTAATLILTDRAEAKAGOLGARVAV 6476
 Oy 20 ----- 19
 Db 6477 DEFGAASGSEADAPGPTGTGTGTSRGYVDAPVEGLRPODLAVYVMTSGTGVPKGV 6536
 Oy 20 ----- 19
 Db 6537 AVTRGVVDVLRDCHMPCRGHERVLLAPAFADVSCYEMVPLVSGTGVVAPRGLDPA 6596
 Oy 20 ----- 19
 Db 6597 AITDLIAHDTAHLNAGFRVVALEAPCEPAGREVLLGGDVSPAAVAVLAHPRI 6656
 Oy 20 ----- YXKRGTPVGR 30
 Db 6657 VLRLVGPTEITLCVTOHEVYAPYKAGSLPVG 6690
 RESULT 4
 titin, cardiac muscle [validated] - human
 H:Alternate names: connectin
 C:Contents: serine/threonine-specific protein kinase (EC 2.7.1.1)
 C:Species: Homo sapiens (man)
 C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Sep-2000
 C:Accession: I38344, I38345; S20898; S20899; S6365; S37393
 R:Label: S.; Kolmerer, B.
 S:Date: 2/0, 293-296, 1995
 A:Title: titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; PMID:96026330; PMID:7569978
 A:Accession: I38344
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
 A:Molecule type: mRNA
 A:Residues: 1-26926 <LMB1>
 A:Cross-references: EMBL:X90568; HID:g1017424; PID:g1017425
 R:Musco, G.; Tzitzalos, C.; Schuck, P.; Pastore, A.
 Biochemistry 34, 553-561, 1995
 A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix
 A:Reference number: I38345; PMID:95119041; PMID:7619249
 A:Accession: I38345
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1977-2014 <MUS>
 A:Cross-references: EMBL:X83270; HID:g602579; PMID:CAA58243.1; PID:g602580
 R:Label: S.; Gabel, M.; Lakey, A.; Trinick, J.
 EMBD J. 11, 1711-1716, 1992
 A:Title: Towards a molecular understanding of titin.
 A:Reference number: S20897; PMID:92258380; PMID:1582406
 A:Accession: S20898
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 13597-14200; T, 14202-14696 <LMB2>
 A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193

A:Accession: S20897
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 16330-16382; S, 16384-16756; F, 16758-16860 <LMB3>
 A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
 A:Accession: S20899
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: P, 22278-22431; R, 22433-22448; G, 22450-22453; Q, 22455-22480; T, 22448
 A:Cross-references: EMBL:X64697; HID:g37190; PIDN:CAA45938.1; PID:g37195
 R:Kolmerer, B.; Olivieri, N.; Wilt, C.C.; Herrmann, B.G.; Labelle, S.
 J. Mol. Biol. 256, 556-563, 1996
 A:Title: Genomic organization of H line titin and its tissue-specific expression in t
 A:Reference number: S63665; PMID:96177761; PMID:8604138
 A:Accession: S63665
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 26729-26825 <KOL>
 A:Cross-references: EMBL:X92412; NID:g1236761
 R:Gabel, M.; Leonard, K.; Labelle, S.
 EMBD J. 12, 3827-3834, 1993
 A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in different
 A:Reference number: S37393; PMID:94008990; PMID:8404852
 A:Accession: S37393
 A:Molecule type: mRNA
 A:Residues: 26831-26926 <GAD>
 R:Improta, S.; Pollitt, A.S.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, February 1996
 A:Reference number: A66736; PDB:1ITI
 A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
 R:Pinhi, H.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, August 1996
 A:Reference number: A66201; PDB:1NCT
 A:Contents: annotation; conformation by (1)H-NMR, residues S, 26059-26155
 C:Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q32
 C:Function:
 A:Description: structural protein forming filaments in striated muscle
 C:Superfamily: titin, fibronectin type III repeat homology; immunoglobulin homology;
 C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; 91
 structural protein
 F:24752-95008/Domains: protein kinase homology <KIH>
 F:84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,40
 98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,1354
 16780,16976,17579,17602,17667,17681,17845,17899,18121,18186,18209,18336,18670,18680
 21900,21935,22295,22495,22627,22897,23024,2318,23883,24012,24177,24290,24447,24642,
 F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experiment

Query Match 42.3% Score 66; DB 1; Length 26926;
 Best Local Similarity 0.1% Pred. Ho. 4.8e+04;
 Matches 22; Conservative 3; Mismatches 6; Indels 24391; Gaps 7;

Oy 1 SRXHX----- 5
 Db 2459 NRTNHSDEGPRYLIVGRVETNCHLSVEKIRIRGLRDLCTFETQVAVVEVLSHGIDVL 2518
 Oy 6 ----- 5
 Db 2519 WNFKEKIPSSKYLLEAHGKLYKLVLMKMDKGYTFYAGENMTSGKLTAVAGALS 2578
 Oy 6 ----- 5
 Db 2579 PLMDQTVASGAVPECEVANDPSKGEMLROCKHLEPLTNHRSBSDCRRLITIAATKLD 2638
 Oy 6 ----- 5
 Db 2639 DIGEYTVANSKTSAKLVNAVKIKKLLKLVETQDAVETVELTPNPKVQWIKNG 2698
 Oy 6 ----- 5

Db 2699 VLESNEKXIAVSGTISLRKNCAYDESVGFRGLGASARLHVEVKKIIRKPKDV 2758
 QY 6 ----- 5
 Db 2759 TALENATVAFEVSVSHDTPVPMKHKSVKIKPSDKHRLVSRKYHKLMLONISPSDAGEY 2818
 QY 6 ----- 5
 Db 2819 TAVVQLECRANKLFEVETLHITKTKRLEVPETKTASFCEVSHENVPMSMLKNGVELEMS 2878
 QY 6 ----- 5
 Db 2879 EKKIVQGLHOLIMNTSTEDSAREYFVCGNDQVSATLTPPIMITSMKLDINAERKD 2938
 QY 6 ----- 5
 Db 2939 TITTEVTVNBSISTKYLKNGVEIKSTDKQMRKLLHSLNINRVHFGAADYTFVAGK 2998
 QY 6 ----- 5
 Db 2999 ATSTATLVEARHIEFRKHIDIVLEKKRAMFCEVSEPDITVQMMKDDQLQITDRIK 3058
 QY 6 ----- 5
 Db 3059 IQEKYVRLLIPSTRMSDAGKYTVAGGNYSTAKLPEGRDVRIRSLIKREVQYIERORA 3118
 QY 6 ----- 5
 Db 3119 VVEFEVNEEDVDAMHTKQGEINFQVQERHKKYVERRIHRMFISETROSDAGEYTVAGR 3178
 QY 6 ----- 5
 Db 3179 NRSVTLVNAPEPQVLOELQPVTVQSGKPARFCAMISGRPOKISWYKEBOLLSTGFK 3238
 QY 6 ----- 5
 Db 3239 CKFLHDSGEYTLLEAPEDAAVYTCARNDYVATTASLSVEVPEVVSPPDEMPYR 3298
 QY 6 ----- 5
 Db 3299 PAITPLQDVTSEGOFPARFCORVSGTDLKVSWSKDKIKRPFRTIQEDTYOLEIA 3358
 QY 6 ----- 5
 Db 3359 EAPDEGTYFVANNAGVVSSTANLSLEAPESILHERIOELEMENKESSTLSAE 3418
 QY 6 ----- 5
 Db 3419 ECLHSAELOSKINETLELSESPVYPTKPSKKGTPITKEVSNADJSMGIVALLSV 3478
 QY 6 ----- 5
 Db 3479 TVIGIPKRIQWFFNGVLLTFSADYKFFVGDGDSHLILFTKLEDEGEYTCMASNDYKGT 3538
 QY 6 ----- 5
 Db 3539 ICSAYIKATNSGEGHKQETESSAAVKSLEKLGPCPHLEKLP IRCAGOLPAIFEXTV 3598
 QY 6 ----- 5
 Db 3599 VGEPAFVTVKRNKQCTSVYITLIHNPNGSGTFIVNDPQREDSGLYICCAENMLGSEST 3658
 QY 6 ----- 5
 Db 3659 CAELLVLEEDTMDTDCAKASTPEADEPQPFPLKSPAVEALDSBOELATFVKDLIK 3718
 QY 6 ----- 5
 Db 3719 AALITEENOQLESTEHIAKANELSOLPLGAQELQSLDOKLTPESTREFLCINGSIHQ 3778
 QY 6 ----- 5
 Db 3779 PLKPSPNLQLOIYOSQKTFSEKGLIMPEDEPTQAVLSDTERTKIPPSANSIHQINSLTVER 3838

QY 6 ----- 5
 Db 3839 IKTLLAPEBGNYPQSSIEPPHMSYLSVAEYLSLKEKTVSDTNRQVYTLQKQENQASAL 3898
 QY 6 ----- 5
 Db 3899 ILSQSLAEGHVESLQSPDVMIQOVNTEPLVPSHSCTEGKILIESANPLENAGQDSAVR 3958
 QY 6 ----- 5
 Db 3959 IEEGSLREPLALEEKQVILKEHSDNVVMPDQIIIESKREPVAIKKVOEGRDLISKE 4018
 QY 6 ----- 5
 Db 4019 SILSGIPEEDRLNKTQICRALQANVASRQPLFBSWLNIEKVEAVNITQEBRHINC 4078
 QY 6 ----- 5
 Db 4079 MYLVTSKSVTEBVTIIEVDVPOMANLKMELDMLCALIYEIDILTAEGPRIOGAKT 4138
 QY 6 ----- 5
 Db 4139 SLOEKDSFSGQKVEPTTEPEVESKYLITSEBVSIFPNQSRVXYLDATPVTAKVASAV 4198
 QY 6 ----- 5
 Db 4199 SDEKQESLKPSEKESSESSEGTAEVATVKIQEABGGLIKEDGPAIHPLVDVSEBD 4258
 QY 6 ----- 5
 Db 4259 IVHLLTSTINAKENVYFENKLVPSDEKFCLODONTYTLVIDKVNTEHDGGEYCEALN 4318
 QY 6 ----- 5
 Db 4319 DSGKTATSAKLYVKKRAAPVIRKPIELEVALGHLAKFTCEIQASPNVRQWFKAGRELY 4378
 QY 6 ----- 5
 Db 4379 ESDKCSIRSKYISLEILRTQVDCGEYTCANESGVSCTATLTVPGEEKVRRL 4438
 QY 6 ----- 5
 Db 4439 LPERKPEKHEVVLKSVLKRPEEBPKVPEPKLEKVPAPVEPPPPVBEZVPTVT 4498
 QY 6 ----- 5
 Db 4499 KREKKIPEPTKVEIKPAILPAPPEPKRPAEAVKTIKPPVBEPTPIAADVTVPVGK 4558
 QY 6 ----- 5
 Db 4559 KAEKAPKEBANPKPGIKVPKTPSPLEABRRRLRPGSGEKPDEAPFTYOIKAVPL 4618
 QY 6 ----- 5
 Db 4619 KFYKEIKDILITSEFYGSSAIFECLYSPSTAITTWAKGDSNRRESPKRPITADCKDKRL 4678
 QY 6 ----- 5
 Db 4679 HIIDVQSDAGEYCYLRLGNKKEKSTAKLVVEELPVREKTLREBVVWVGQPLYLSC 4738
 QY 6 ----- 5
 Db 4739 LNERDVYWRKDKIYERKGRITVPGVIGMRALTINDADDTAGTYTVVENANNLECS 4798
 QY 6 ----- 5
 Db 4799 SCVKYVAVJLDMLYKPIRDQHVKPKGTAFACDIANDTINIKFKGYDILPAEPNDKTEI 4858
 QY 6 ----- 5
 Db 4859 LRDGNHLTYLIKNNAMPEDIAEVAVEIEGKRYPAKTLGGERVELLKPIDVTVYERKSAS 4918

QY 6 ----- 5
Db 4919 FBAEISADIPGQWIKGELLRSPICEIAEGGKRLILAKVLDQAGEVILQALNALT 4978
QY 6 ----- 5
Db 4979 TAILTYKEILDRAVPLKDYTPERROARPECVLTREANYNSKPDIIKSSDKFDIAD 5038
QY 6 ----- 5
Db 5039 GKHLIIVINDSOFDEBQVTAEBGKTSARLFVTGIRLKFNSPLEDOQVKEGETATFVC 5098
QY 6 ----- 5
Db 5099 ELSHEKHEVWFKNDANLMTSRTVLISSEKTRHEKKEVTLDDISOIKAQVNEISSTAQ 5158
QY 6 ----- 5
Db 5159 LKVLADPYTYVKLHOKTANERKDEITLKCHVSKDVPYKTKOGEELVPSPKYSIKADGLR 5218
QY 6 ----- 5
Db 5219 RIUKIKRADLKDGEVYCCGDTKANTVTEARLIEVEKPLYGEVEVGETAHEIELS 5278
QY 6 ----- 5
Db 5279 EPPVHGQWIKLGPPLIASPDEIIEGKHHILHNCQJLMTGFEVSFOAANAKSAANLKY 5338
QY 6 ----- 5
Db 5339 KELPLIITPLSDVKYERKEDEAKFECEVSRPEPTFRMLKQTOEITIGDDRELKDGKHS 5398
QY 6 ----- 5
Db 5399 MYKSAFEBDEAKYMERAEKDHTSGKLLIEGIRLKEPLTKDYTAKEKESAVTELSHD 5458
QY 6 ----- 5
Db 5459 NIKVWFKNDQRLHTTSVSMQDEKTNSTIFKOLSIDTSTQIRVEAMGSSAKLTVLE 5518
QY 6 ----- 5
Db 5519 GDPYFGKLDQYGEKDEVILCEISKADAPYKFRDGEIKRPSNNAVKTGKKRMLI 5578
QY 6 ----- 5
Db 5579 LKRALSDIGQYTCDSGTDKTSGLDIEDREIKLVPLSHSEVAMEETAFETEISED 5638
QY 6 ----- 5
Db 5639 HANMKLGBALLQTPDCEIKESBKHSILVJLHNCRLDQGTGVFOAANVSSAHLRVPKY 5698
QY 6 ----- 5
Db 5699 IGLRLPKDYTAGETATFDCELSEDIPEVWYLGKLEPSDKVYPRSEGVHLLILR 5758
QY 6 ----- 5
Db 5759 DVKLEAGEVOLTAQDKFTANLFEKPPVEFTKPLEDQVEBGAVALCECEVSRENKY 5818
QY 6 ----- 5
Db 5819 KFRKNGEILKSKKYEIVADRYRKLYIHDCTEEDIKTTCDAKDKTSCNLNVPPHNE 5878
QY 6 ----- 5
Db 5879 FLRLPLDQVREKEMAFECESLRENAKYWFKDGAIEKKGYDILSKGAVILIVNKC 5938
QY 6 ----- 5
Db 5939 LIADBAFYSCEVNTARTSGMLTVLEBAVTKMLANIESETPTIKLVCEVSKFGAEVIM 5998
QY 6 ----- 5

Db 5999 YKGBEIIETGAYEILTEGRKRILVIONAHLBEDAGNMCRLPSSRTDGVKVEHIAEFI 6058
QY 6 ----- 5
Db 6059 SKPQNLIEGEKAFCVSISKESFPVOKKDDKTLSEGDKYDIADGKRRLVYKATL 6118
QY 6 ----- 5
Db 6119 QDMGYVWVGAAARAHLTVIEKLRIVPLKDIRKQGEVVENCVNTEGAKAKMER 6178
QY 6 ----- 5
Db 6179 EBAIFDSKYIILQDLVYTLIRDAHLDDQANYVSLTNIHGENYKSAANLIVEEDLR 6238
QY 6 ----- 5
Db 6239 IVEPLKDIETMEKSVTYTCKAVNRLNVLTKYKNGEVPFDNRYSYVDKTKHMLTIKDG 6298
QY 6 ----- 5
Db 6299 GPPDBGEYLVTAGDQSVAEILLIETAPTEVEHLEDQVTERPDVAFSCQLSREKANYK 6358
QY 6 ----- 5
Db 6359 YNRGREIKGKKYKEKDGSIHRLIIKDGLDCEYAGVEDRKSARLFEVEIPEI 6418
QY 6 ----- 5
Db 6419 RPPDILFAGADVTFELANRDKVEYQMLRNMYVVOGDHQAISEKIHRLQICDKP 6478
QY 6 ----- 5
Db 6479 RDQGEYPLAKDKKARAKLLELAAPKIKYADQDLVVDGKPLTWVVPYDAVKAEMFK 6538
QY 6 ----- 5
Db 6539 ENEPLSTIITPAOTSRILEAKKGDKGRYKIVLQNRKAGELNLKYIDVPGPYRN 6598
QY 6 ----- 5
Db 6599 LEVTEFDOEVSILAMEPLTDGSKIIIGYVERBDIKKRTVLAIDRAESCEFTYTGLOK 6658
QY 6 ----- 5
Db 6659 GGEVYLFRYSANRYGTGEPEFDNPVEARSKYDVPPLNYTIDVNRFGVSLTWEPPE 6718
QY 6 ----- 5
Db 6719 YDGAELITNYLELBDKTSIKMDTAMTYBAEDLSATVDVVEGQESFRVBAQNRIGVK 6778
QY 6 ----- 5
Db 6779 PSAAFPVAVADPIERPSPVNLJSSDQTSQSVQAKMEPLKDGSPILGTYIERCEBG 6838
QY 6 ----- 5
Db 6839 DMWIKONKALVPELTYKYTGLEKGNKLYRYSAENKAGVSPSEILGLPLADAEVEPTM 6898
QY 6 ----- 5
Db 6899 DLSAFKQGLEIVVPPNLTILVPSGTGPPPLATWCGRVLETDGRVKKTLSSAVALVIS 6958
QY 6 ----- 5
Db 6959 PERSDKGIYTLKLEBNVKTISGEIDVNVIAVSPAKELKGDITKDSVHLTWEPDDG 7018
QY 6 ----- 5
Db 7019 GSPITGVYERKREVSRTKTYKVMDFVTLDEFTVPLDVGKELYFKVCARNKCGRGPAYV 7078
QY 6 ----- 5

```

Db 7079 DEPVNKSPTATVDPDPENYKMDRTANSIFLTWDPKNDGSGRIKGYIYERCPGSDKW 7138
QY 6 ----- 5
Db 7139 ACCEPVAETKMEVTOGLEBGRWAVYRVKTLNRQGASKPSNPTEIQAVDTQEAPELFLDV 7198
QY 6 ----- 5
Db 7199 LLAGLVKAGTIELPATVTKPEPKITWTKADMILKODRITFENVPKSTVTVDSKR 7258
QY 6 ----- 5
Db 7259 SPTGTYIIANVAVCGRATAVVEVAVLDRGPPAADITDVYNESCLLTWNEPRDDGSKI 7318
QY 6 ----- 5
Db 7319 TNYVERRATDSBVMHKLSSTVKDYNFKATILIPKREYIFRYAENMTGAGHPQASPT 7378
QY 6 ----- 5
Db 7379 ANQOPDPGPPTLEPSDITNDATVLTWCEPDGSGPIYGYVERLDPDIDKAVRCNM 7438
QY 6 ----- 5
Db 7439 PYKDTTYRVKGLTNKKTYFRVLAENLAGPKPSKSTEPILIKDIDPPMPGKFTYKV 7498
QY 6 ----- 5
Db 7499 GKTSVRLNTRKPEHDGAKIESYIEMLKTGTDENVRVAEGVPTQHLPLGMEGQESF 7558
QY 6 ----- 5
Db 7559 RYRAVNKAGESPSBSPVLQREKLYPSPRMLEVINITKNTADLKTVPEKDGSP 7618
QY 6 ----- 5
Db 7619 TTYIYERKRVARKQVDTTYTKDKCYVTPLEBGLYFRVAENAIQSDYTEIEDSV 7678
QY 6 ----- 5
Db 7679 LAKDTFTYGPPLYALAVDVTRHVDLKEWPKKNDGSRIGRYIEKKEBLGTRVAKAK 7738
QY 6 ----- 5
Db 7739 TAGPDCNFRVTDVIBGTEYQVRAENAGVGPSEPTIELSIEDPTSPSPPLDLHYTD 7798
QY 6 ----- 5
Db 7799 AGRKHLAIAKPEKNGSGPIIGIYVENCYGTETKMRVNSRPIDLKFVEBGVPEKE 7858
QY 6 ----- 5
Db 7859 YVLRVAVNAIGVSESEISENVNAKDPCKPTIDLETHDIVIGEKLSIPVPRAPV 7918
QY 6 ----- 5
Db 7919 PTVSHKDGKEVAKSDRLTKNDHISAMLEVPKSVRADAGIYTTIENKLSATASINVK 7978
QY 6 ----- 5
Db 7979 VIGLPGCKDIASDITNSSCKLTWEPPEFGGTPIILHYLERRAGRRTYIIPVAGSK 8038
QY 6 ----- 5
Db 8039 LSWTKDLIPNGEYFRVAVKAVGSGEYIELKNVIAQDPQPPDPVDVAVHPTAA 8098
QY 6 ----- 5
Db 8099 KWTYKPLYDGSRIKGYITTEKIAKGERKRCNEHLPIITYAKGLEKQYQFVR 8158
QY 6 ----- 5
Db 8159 AENAGISEPSRATPTTAVDIDAPKYLRTSLEVRGDELTALDASISGSPYPTTWIK 8218

```

```

QY 6 ----- 5
Db 8219 DENIVPEIKKNAAPLYRRKRGVQEEPEFVLPYQRLSIDNSKKGESOLRVROSLRD 8278
QY 6 ----- 5
Db 8279 HGLYIYENDHGIAKAPCTVSLDPPGPINEVFEDIRKTSVLAGWPEPLDGGSEIN 8338
QY 6 ----- 5
Db 8339 YTLEKDKTPEDEMYVTYTGILRECKISTYKLEGEYLFVRAENRFGGPPCVSKPLY 8398
QY 6 ----- 5
Db 8399 AKDPGPPDAPDKPIVEDYTSNSMLVKNNEPKDNGSPILGWLKREVNSTHMSRYNKG 8458
QY 6 ----- 5
Db 8459 LNLKANVDILBGLTYFVRYCAENAGPKFSPPSPDKTAHDPISPPGPIPRVIDIS 8518
QY 6 ----- 5
Db 8519 TTLELEPPAFNGGGEIYGTVDKQLVGYNKSRECTEKIKYQYTYEIRGADYKLR 8578
QY 6 ----- 5
Db 8579 VSAVNAAGBGPGEPTOPVTAEPQEPVAVELDVSXGCIQIMAGKTLRIPAVVTGRPVPT 8638
QY 6 ----- 5
Db 8639 KWTKEGEELDKRVAVIDNVGKSELINDALRKHGRVYITAINSGSKFAAARVEVD 8698
QY 6 ----- 5
Db 8699 VPGVYDLKPYVYTKRACCLNMSDPEDDGSEITGFIIEKAKAKHTWQPIETERSKD 8758
QY 6 ----- 5
Db 8759 ITGLEGEYKFRVIAKNGCGPVEIGPILAVDPLGPTSPERLTYTERQSTITLDM 8818
QY 6 ----- 5
Db 8819 KEPSNGSGSPIQYIIEKRHRDKPDERVNRCLPTTSFLENLDRHOMYEFVKAANEL 8878
QY 6 ----- 5
Db 8879 GESEPSLPLNAVIODDEVPTIKRLSYRGDTIKVAGEPVHIPADVTGLPMKIEMSKN 8938
QY 6 ----- 5
Db 8939 ETVIEKPTDALQITKEVSRSEAKTELSIPKAVREDKGYTYTASNRLGSVFNHVEY 8998
QY 6 ----- 5
Db 8999 DRSPPRNLAVTDIKASCYLTWDAPLDNGSEITHYVIDKRDASRKAWEVEVNTAVE 9058
QY 6 ----- 5
Db 9059 KRIGIKLIPNGYEFVRAVNTYGISDECKSDKVYIODEYRLPGPPGPKYLAIRKGSN 9118
QY 6 ----- 5
Db 9119 LVSVPPLDNGGSPITGYWLEKREBSPYMSRVSRAPITFVGLKGYEFNVRPLEGVKQ 9178
QY 6 ----- 5
Db 9179 FRANAANAAGIPSEPSDEVAGDPIFPGPSPCEYKDKITSISLGMKPPAKDGSGP 9238
QY 6 ----- 5
Db 9239 IKGYIYMEGEGTTDMKRVNEPKLITTCBQVVPNKIEKRYFRVAVNAEGESEPSDT 9298

```

QY 6 ----- 5
Db 9299 TGEIPADIDOEPEVFLIDGADCIACKAGSQRIPAVIKGRPTPKSSWEDGKAKKAMK 9358
QY 6 ----- 5
Db 9359 DGVHIDIEDAOLETAENSSVILIECKRSHTGKSTIAKKNAGOKTANCRCRVKNDVGGPP 9418
QY 6 ----- 5
Db 9419 KDLKVSIDITRGSCHLSWKMPPDDGDRKIGYIEKRTLDGKAMTKVNPDDGSTTFVVPDL 9478
QY 6 ----- 5
Db 9479 LSEGOYFFRVBAENRFGIGPVEET IORTADPIPPDPPIKIKIGLITKNTVHLSWKP 9538
QY 6 ----- 5
Db 9539 KNDGSPVTHIYECIAMDPGTGTCKEANKRCNKRDVELOFTVEDLYEGGEYEFVKAAN 9598
QY 6 ----- 5
Db 9599 AAGVSKSATVPCDQORPMPISIDKEFMEVEEGTNVNIKIGVPEPTLTFWKAP 9658
QY 6 ----- 5
Db 9659 KKPDKNEPVLYDTHNKLIVDDCTIYIPQSRSDTGLYTITAVNNLGTASKBHRNLVIG 9718
QY 6 ----- 5
Db 9719 RPPGPVPIKESVSADQMTLSWPPKDDGSKITNTYIEKREANKRTVHVSSEPECT 9778
QY 6 ----- 5
Db 9779 YTIKLEGHBYERMAONKTGIGELDSEPETARMLFVPGAPDKPTVSVTRNSMTV 9838
QY 6 ----- 5
Db 9839 NMEPEIDGSPVTGYLEKDOTTSKWKRVNRPDKAMTLGYSYKVTGLIESDYQFRV 9898
QY 6 ----- 5
Db 9899 YAINAGVPSALSPDAPATARDPIAPGPPEPKVTDWTKSSADLEMSPLKDGSKVTGY 9958
QY 6 ----- 5
Db 9959 IVEYKECKEEMKEKADKEYRGTKLVVTGLKESAFYKFRVSAVNIAGIGEGEVTDIEM 10018
QY 6 ----- 5
Db 10019 KDRLVSPDLQDASVRDRIVVHAGVIRITAYVSGRPPTVTNNMNERTLPOENITETTA 10078
QY 6 ----- 5
Db 10079 ISSSMVKNQCRSHQGYSLAKNKEGERKKTIVDVLDVPGVGPPLAHLNLINESCKL 10138
QY 6 ----- 5
Db 10139 TWSPEDDGSPITNTYIERESDRAMPVTVTRONATVOGLIQKAVFRIAENS 10198
QY 6 ----- 5
Db 10199 IGMGPVETSEALVIREPITVBERPDLKVEYTKNTVTITNPPKXIDGSEIINNYVLS 10258
QY 6 ----- 5
Db 10259 RLIGTEKHFVTDNLSRKTYTKGLESDPTVEYRSVAVNIYGKSPFCTKPTCKDEL 10318
QY 6 ----- 5
Db 10319 APPTLMLDFDKLTIRVGAEPALIGRYSGKPKRVSNFKDEADVLEDDRTHTIKTTPATIA 10378
QY 6 ----- 5

Db 10379 LEKIAKRSDSKXYCVYVENSIGSRKFGCQVNVYDHPGPVGPVSRDEVTXDNVYSKMP 10438
QY 6 ----- 5
Db 10439 PLDDGSKITNTYIIEKKEVKDWMPEVTSASAKTCKVSKLEGKDYIFRIHAENLYGIS 10498
QY 6 ----- 5
Db 10499 DPLVDSMKAKDRFVYDAPDOPITVEYTKDSALVYMNKPFDHGGKFTNTYIIEKREYMSK 10558
QY 6 ----- 5
Db 10559 RMARVTKDPIHPYTKFVBDLEGCQYEFVSAENEIGIDSPSPSKPVFANKDPLAKDSP 10618
QY 6 ----- 5
Db 10619 PVNPEALDTTCNSVDLTWOPPRHDGSKILGYIVEYQKYGDEMRANHTPSCPEYKX 10678
QY 6 ----- 5
Db 10679 VTGLRDCQYTKFRVLAVNAGSDPANVPEVLYVDRLEPPELITDANKARQHTKVDGT 10738
QY 6 ----- 5
Db 10739 LRLSAILKGVPPEPKYWKEDRDAPTKRIDVTVGSKLEIENNAHEDGISTVENPA 10798
QY 6 ----- 5
Db 10799 GSKTVSVKVLVDKPGPPRDLVESEIRDSCTLTKEPLDDGSAVITNTVERRDVASNO 10858
QY 6 ----- 5
Db 10859 WSPLSATSKKSHFAKLENGOYLFRVAENQYGRPEVETPKIKALDPLHPGPPEKD 10918
QY 6 ----- 5
Db 10919 LHHVDVKTESLVNWKPRDDGSPITGYLVEYOEGTODMIFKFTVNLCEVVTGLQOG 10978
QY 6 ----- 5
Db 10979 KTYRFRVAKENIVGLPPTTIPICQEKVPPSVELDYKLIBGLVYVAGTVERPALIR 11038
QY 6 ----- 5
Db 11039 GVPVPAKWTTDGSEIKTDEHYTERTDNFSSVLIKNCLRDTGEYQITVSNAAGSKTVA 11098
QY 6 ----- 5
Db 11099 VMLTVLDVPGPPTGINIIDVTPEHNTISWOPKDDGSPVINTYIEKODTRKDTWGYVS 11158
QY 6 ----- 5
Db 11159 SGSSKTKIKIPIHLOGCEYFRVBAENKIGVPLDSTPTVAKKRFSPSPGKPVWTDI 11218
QY 6 ----- 5
Db 11219 TENAATVSWTLPKSDGSPITGYIYMERREVTGKVRVAKTPIADCKFVGTGYEGNTVEF 11278
QY 6 ----- 5
Db 11279 RVFAENLAGLSKSPSSDPAKACRPIDPPPIINPLKDKSRETAOLWTKPLSDGSGPI 11338
QY 6 ----- 5
Db 11339 LGYVEEQKPGTAQNNRINKDELINQCAFVRVGLLEGNEFRKIAANIVGEGPRELAE 11398
QY 6 ----- 5
Db 11399 SVTANDILPPEVELDVACRDVYIVRVGQITRIILARVGRPEPDIWTKEGVLVREKRV 11458
QY 6 ----- 5

Db 11459 DL1ODLPEVELQ1KEAVADHGKYLIIISAKNSSGHAQSAIWNVDPRPCQNLKVTNWK 11518
QY 6 ----- 5
Db 11519 ENCTISWENPLDNGSEITNFIVEYRKNOKGWSIVASDVTKRLIKANLLANNETTYRVC 11578
QY 6 ----- 5
Db 11579 AENKVGCP1IETKPIIAPIDRPGEPENH1ADKGKFFVYLAKWRPDYDGGSPMLSY 11638
QY 6 ----- 5
Db 11639 HVEBRLKSGDDMEYHKGSIKETHYNDRCVENQ1YEFVQTKNEGSDPVKTEEYVVK 11698
QY 6 ----- 5
Db 11699 EDLOKPVLDLKSGLVLYKAGDTIRLBAVGRKPPPEVAMTKDKDANDLTRSPVKIDTR 11758
QY 6 ----- 5
Db 11759 ADSNFSLTKAKRSDGKTYVATNTAGSFVAATVNVLDKPGVRMLKIYVSSDCYV 11818
QY 6 ----- 5
Db 11819 CWDPEDDGGEIQNYLLEKCEKRWMTYSATVLPFTTVTRLIGNEXIFRVAENK 11878
QY 6 ----- 5
Db 11879 IGTGPTESKPVIAKTKYDKGRDPPEVTVKSEEMTVVANNPPEYDGGSLTGYFLEKK 11938
QY 6 ----- 5
Db 11939 EKHSRNVVANKSAIPERRMKVQNL1PDHEYQFRKANEIGEPSP1PSRPVAKDP1E 11998
QY 6 ----- 5
Db 11999 PPGPTNFRVYDTKHSITLGHGKFTYDGNP1IGYVEMKPK1ADASPDEGMRCAAAA 12058
QY 6 ----- 5
Db 12059 QLVKREFTVTSLENDQEFEPVCAQONQVIGRPAELKKA1KPKRILBPELIDLASMRKL 12118
QY 6 ----- 5
Db 12119 V1VRAGCP1RUFALVGRPAKVTMRKVGDVNVKGOVDLVDMAPLV1PNSTRDSGR 12178
QY 6 ----- 5
Db 12179 YSLTVNPAKEKAVVNVNVLDTGPPVSDLVSDVTKTSCHVSNAPBENDGSGVTHYV 12238
QY 6 ----- 5
Db 12239 EKREADKRWTSVPEVEKTSFHVNLVPGNEYFRVTAVNEGPGVTDVDPKVLASDP 12298
QY 6 ----- 5
Db 12299 LSEDPKRLBATEMTKNSATLAMLPLLDGAKIDG11SYREBEPADRMTESVYVKD 12358
QY 6 ----- 5
Db 12359 LSLVVTGLKEGKKYKFRVAANNAVSLPREAGVTEAKBOLLPKILMPQITIKAGK 12418
QY 6 ----- 5
Db 12419 LR1EAHVYKGPPLCKMKKGEDEVYSSHLAVHKADSSS1LIRDVTRKDSGYSLMAEN 12478
QY 6 ----- 5
Db 12479 SSGTDTOKIKVYVADBPPOPPEDISDIDADAGSLSHW1PLEDGGSH1NYIYKCDVS 12538
QY 6 ----- 5
Db 12539 RGDVYTA1ASVTKSCRGKLLPGQETIFRVAENRFG1SEPLTSPKVVAPFP1GVPSEP 12598

QY 6 ----- 5
Db 12599 KNAKVTYKNDCLFYAMDPRDSDGSP1IGYLEKERNSLIRVANDTVRSTYPCAG 12658
QY 6 ----- 5
Db 12659 IVEGLETSFRYALNKAQSSPSKPLEYV1ARMPVDPGKPEVIDVTKSTVSLIWARPKH 12718
QY 6 ----- 5
Db 12719 DGGSK1IGYFEACKLPGDKWRCNTAPHOLPOEETATGLEKAQYQFRAIARTAVNIS 12778
QY 6 ----- 5
Db 12779 PPSRSDPVT1LAENVPRR1DLSVAMKSL1YKAGINVC1DATVFGKPM1YSSKKGDTL 12838
QY 6 ----- 5
Db 12839 LKPAEGIKRANORNLCTLELFEVNRDSDGYT1TAENSSGSKSATIKLVLDKPGPPASV 12898
QY 6 ----- 5
Db 12899 KINKKTSDRAMLSWEP1LEDGSGSEITNY1VDRKETS1RPMNAOVSA1VP1TSCVEK1EG 12958
QY 6 ----- 5
Db 12959 HEYQFRICAKENKYGVDPVFTEPA1AKNPDYDPPGRODPV1SNTKDHMTVSMKPPADGG 13018
QY 6 ----- 5
Db 13019 GSPITGYLLEKRETOAVNWKVNRKPIERTLKATGLQGTETEFRVTA1NKAQPGKPSD 13078
QY 6 ----- 5
Db 13079 ASKAARBDPQPPAPPA1PKYVDTTRSSV1SMKPRATDGGSP1IGYLYEKADSDNM 13138
QY 6 ----- 5
Db 13139 VRCNLPONLOKTRFVTLGMEOTOYQFRVAVANK1GSDPDPKHPKDL1LPREGH 13198
QY 6 ----- 5
Db 13199 DADLAKFTL1LRAGVTNRLIYVAKGRPPK1TWSKPNVNLDR1GLD1STDFDTLRCEN 13258
QY 6 ----- 5
Db 13259 VNKYDAGKYL1TLENSGKRETT1YKVLDTGPP1NVTYKELSKDSAVTWEP1IDGG 13318
QY 6 ----- 5
Db 13319 SPL1NTVVOKRDAERKSWSTVTECSKTSFRVNP1NEEGKSYFFRVAENEG1GDPGETR 13378
QY 6 ----- 5
Db 13379 DAVKASQTPQPVYDLKVRSGVSKSSCS1GKAKKPHSDGSKR1IGYVDFLEENKQVRZMS 13438
QY 6 ----- 5
Db 13439 LSLQYSAKDL1BEGKETF1RVAENEGEGTPE1FYVARDVYAPDL1DKG1DLCYLAK 13498
QY 6 ----- 5
Db 13499 ENSNFR1K1PIKGRPA1SVSMKKGEDPLATDTRVSVSSAVNTT1LVYDQKSDAGKTY1 13558
QY 6 ----- 5
Db 13559 TLKNAVGTKEGT1SIK1VKGPG1PTGP1KFDVYLAAMTLAKRPA1PRDDGSEITNY1LEK 13618
QY 6 ----- 5
Db 13619 RDSVNNKVVYCASAVQKTFPRVTR1LHGBMEYTFRVAENKYGVEGLKSEP1VAARHPDV 13678

```

QY 6 ----- 5
Db 13679 PDAPPPIVDVRHDSVSLWTDPKKGTGPIGYHLEFKERNSLMKRANKTPIRHDRF 13738
QY 6 ----- 5
Db 13739 KVTGLTBSLEYEFKVMANINLAGVCKPSLPSBPVVALDPIDPPGKPEVINITRNSVTLLMT 13798
QY 6 ----- 5
Db 13799 EPKYDGGKILGYIYERKDLPSKSMEMANNVNEPCAFVTVLVEGKYEERLRANKTAG 13858
QY 6 ----- 5
Db 13859 AISAPSESTETIICKDEYEAPTYVLDPTIKDGLIKAGDITVINALSLCKPLPKSSWSK 13918
QY 6 ----- 5
Db 13919 AGNDIRPSDITQITSTPISMLIKYATRKDAGEYITATNPCTKVEHKVTVLDVPCP 13978
QY 6 ----- 5
Db 13979 PGVEISNVAEKATILWTEPLEDGGSPIKSYILEKRETSRLMTVISEDIGCRHVATK 14038
QY 6 ----- 5
Db 14039 LIQNEIYIFRVAVNHKGGEVQSEPVKAVDRNGPPGPEKPEVSNVTNTATVSKMRP 14098
QY 6 ----- 5
Db 14099 VDDGSEITGYHVERREKSLRWVRAIKTVSDLRCKVTGLOGSGTYEFERVAENRAGIG 14158
QY 6 ----- 5
Db 14159 PPSASDSVLKDAVPPGPPSPHYDTTKKASLWMPKHGDLGELTGYVEHQVQ 14218
QY 6 ----- 5
Db 14219 DEAMIKDTGTALRITQFVVPDQTEKYNFRISAINDAGVGPVAVIPEVEYEREMADP 14278
QY 6 ----- 5
Db 14279 FELDAELRTIIVRAGISIRIFPIKGRPAPEVWTKNONLNKRANIENBESFTLLIP 14338
QY 6 ----- 5
Db 14339 ECKRDTGKFWTLENPAKKGSEFVAVRVLDTARPSPOLRPMDDITKDSVTLHMDLPLIDG 14398
QY 6 ----- 5
Db 14399 GSRITNYVEKRENRKSYSTATTKCHKCYKVTGLSEGCETFRVMAENEGIGEPTEF 14458
QY 6 ----- 5
Db 14459 TEFVKASEAPSPDSINIMDTKSTVSLWMPKPKHDGSGNITGYIEAORKGSDQWTHIT 14518
QY 6 ----- 5
Db 14519 TVYGLCEVVRNLTGSEETFOVMAVNSAGRSAPRESRPVYKEQMLPELDLNGIYQKLV 14578
QY 6 ----- 5
Db 14579 IAKAGDNKEIPLVLRPKPTVWKKGDQOLKQTORVNETTATSTIILINIECVMSDSCP 14638
QY 6 ----- 5
Db 14639 YPLTARNIVGEGDVITIGVHDIPGPTGPIKFDVSSDFVFSMDPPENDGVPISNTV 14698
QY 6 ----- 5
Db 14699 VEMKQDSTTWELATTVIRITTKARLTGLTGLYQFRVAQNRKYGVPGITSAMIVANP 14758
QY 6 ----- 5

```

```

Db 14759 FKVPBPPTPOVAVTKDSMTISWHEPLSDGSPILGYHVEKERNGLMQIVSKALVPG 14818
QY 6 ----- 5
Db 14819 NIFKSSGLIDGIAVEERVAENNAKSKPSKSEPMALDPIDPPGKPVPLNITRHTVL 14878
QY 6 ----- 5
Db 14879 KNAKPEYTGFKITSIYERKDLPNGRKLIKAFNSINLENEFTVSLTEDATEFRVIAKN 14938
QY 6 ----- 5
Db 14939 AAGALSPSEPSDATCRDDVEAPKIKVDKFKDVTILKAGAPRLADVSGRPPPTMEW 14998
QY 6 ----- 5
Db 14999 SKDGKELEGAKLEIKIAOFSTNLVKNKSTRDSGATYLTATNPGFAKHIFNVKVLDRP 15058
QY 6 ----- 5
Db 15059 GPPEGPLAVTEVTSKCVLSMFPPLDDGAKIDHYIVOKRETSRLMTNVASEVQVTKLK 15118
QY 6 ----- 5
Db 15119 VTKLKNEIYFRVAVNKGVEPLESEPVLAVPYGPDPKPEVYTIKDSMYCM 15178
QY 6 ----- 5
Db 15179 GHPSDGSSEIINITYERDKAQORIKCNKTLIDLRYKVSGLTEGHEFRMAENAA 15238
QY 6 ----- 5
Db 15239 GISAPSPSPFYKADYERKPPGPNRYLDTSSRSISIANKPKTYDGSSEITGVNZA 15298
QY 6 ----- 5
Db 15299 LPEDEMOITVPPAGIKATSYITIGLTENOBYKIRIYAMNSEGLPEPALVGPRAEDRM 15358
QY 6 ----- 5
Db 15359 LPELELDADLKVYTIACCTLRIFVPIKGRPDEVRANDHGESLDAKASIESASTYL 15418
QY 6 ----- 5
Db 15419 LTVGVNREDSGKIITVENSSEKSAFVNVAYLDPGPQDLKXEVTKTSVTLWMDP 15478
QY 6 ----- 5
Db 15479 LLDGSKIKNYIYERKESTRKAYSTVAATNCHTKSMKVDQLOGSYYFRVLAENEYGIGL 15538
QY 6 ----- 5
Db 15539 PAETAESVAKASERPLPBGKITLMDVTRNSVLSWEKPEHDSNLTGYIVEMQTKGSDKW 15598
QY 6 ----- 5
Db 15599 ATCATVKTLEATTIGLLOGEESFPRVAQNEKGISDPQULSVPIAKDLVIPAFAKILFN 15658
QY 6 ----- 5
Db 15659 TETVLAGDLKVDVPIGRPTPAVTWMDNVLKQTRVNAESTENSLTINDACREDEY 15718
QY 6 ----- 5
Db 15719 GHVVKLINSAGEALIELNVIYLDKPGPTPVKMDVETADSTLSMGPPTIDGSSILNM 15778
QY 6 ----- 5
Db 15779 YIVERDITSTTWQIVSATVARTIKACRLTKGCEYOFRIAENRYGKSTYLNSEPTVAQ 15838
QY 6 ----- 5

```

```

Db 15839 YPEKPPGPPTPVVTLSSRDSHEVQWNEPISDGSRVIGYHLERKERNLIMVKNKPTI 15898
QY 6 ----- 5
Db 15899 PQTKEKTGLEGEVEFEFRVSAENIVIGKPSKVSCTYAKDPCDPPRREALITVTRNSV 15958
QY 6 ----- 5
Db 15959 TLOMKRFTYDGSKITGYIEKKELPEGRMKASFTNIIDTFEYVGLVEDHREFFRYLA 16018
QY 6 ----- 5
Db 16019 RNAAGVSESESTGALTANDEVDPPRISMDPKDTIVYRAGESEFKVDADIGKPLIPTI 16078
QY 6 ----- 5
Db 16079 QMIGDDELSTABLEIKSIDETSLSKDAVRDGSNTILAKKNVAGERSYVNVKVID 16138
QY 6 ----- 5
Db 16139 RPPPEPVPVVISGVTAEKCTLAFRPLQDGSIIINIVERRERSRLVTVVDANYQTLS 16198
QY 6 ----- 5
Db 16199 CKVTKLEGNETFRIMAANKVGVGEPLSEPPYAKNPVVPDAPKAPBVTITVKDSMTV 16258
QY 6 ----- 5
Db 16259 VMERPADGGSSEILGYILEKNDKEGIMTRCKRLIGELRLVYGLTENHDEPRVSAEN 16318
QY 6 ----- 5
Db 16319 AAGISEPSPSAYOKACDPYKPGPPNNPKVIDITRSSVELSWSKPIYDGSCEIOGYIVE 16378
QY 6 ----- 5
Db 16379 KCDVNVGHWYCTPPTGINKTINIEVKELLEHEYNFRICAINRAGVGEHADVPGPIVEE 16438
QY 6 ----- 5
Db 16439 KLEAPDIDIDLELRKITINIRAGSRLRFLVPIKGRPTPEVKWGVDEIRDAIITDVTSE 16498
QY 6 ----- 5
Db 16499 TSLVDNVRYSGRYTLTLENSGTSKSAFYTVRLDTPSPVNLKYTEITTKDSVSITWE 16558
QY 6 ----- 5
Db 16559 PPLDGSKIKNIYERKEATRKSYAAVVTINCHKNSWKIDQLQSGSYPRVTANENYGI 16618
QY 6 ----- 5
Db 16619 GLPACTADPIFAVPOPPGKITVDVTRNSVSLSWTRPEHDGSKIIOYIVEMQAHSE 16678
QY 6 ----- 5
Db 16679 KMSCARVKSIOAVITTNLTQGEETLPRVAVANEKGRSDPRLANPIYAKDLVIEPDVKA 16738
QY 6 ----- 5
Db 16739 FSSSYQVGDLMKEVPISGRKPPIITWKDGLPLKQTRINVTDSIDLTLISIKETNKD 16798
QY 6 ----- 5
Db 16799 DGGOGITVANVGOQTASIEIVLDRKDPKGVKTDVSAESITLSMNPPLYTGCGOI 16858
QY 6 ----- 5
Db 16859 TNYIQRKDDTTTWWVDSATVAATITLKVTKLGTIEQPIFAENRYGGSFALSDPIV 16918
QY 6 ----- 5
Db 16919 AQYPRKEPBPPTPATAISKDSMWIQWHEPVNNGGSPVIGYHLEKERNSLMTKVNT 16978

```

```

QY 6 ----- 5
Db 16979 IIHDTQFKQNLSEIEYEFRYAENIVGVGASNSCYVARDPCDPPTEPIHVKRN 17038
QY 6 ----- 5
Db 17039 EYTLQWTRPVYDGSMTIGYIEKRLDPGRMKASFTNVIETQFTVSGLTEDQRYEPRV 17098
QY 6 ----- 5
Db 17099 IAKNAAGAIKPSDSTGPIYAKDEVEIPIRISMDPKFDTIVYVAGETPLBVDYHGRPL 17158
QY 6 ----- 5
Db 17159 TLEWLRGDEIESARCEIKNTDFKALLIVKDAIRIDGGYILRASNVAGSKSEFPVNVK 17218
QY 6 ----- 5
Db 17219 LDRPPEPVPVQVGTYSKSLTWSPLQDGSISHTYVEKRETSRLANTVYVASEVYT 17278
QY 6 ----- 5
Db 17279 NSIKVTKLEGNETFRIMAANKVGVGEPLSEAPVLMKNPVLPGPPKSLVYTNIAKDSM 17338
QY 6 ----- 5
Db 17339 TYCWNNPDSGSEIIGYIEKRRDSGIRWIKONRRITDILRLVYGLTENDEYEFVSA 17398
QY 6 ----- 5
Db 17399 ENAGVGEPSPATVYKACDPYFKPGPPTNAHITVTTKNSITLANGKPIYDGSSEITIGY 17458
QY 6 ----- 5
Db 17459 VEICKADEHEWQIVIPQGLRVYRFEISKLEHQBKIRVCAIKNVGLGEANTVPGTKP 17518
QY 6 ----- 5
Db 17519 EDLEAPELDSELRKGIYVAGSARIIHPKGRPMPEITWSHEGEFTDKVOIEKV 17578
QY 6 ----- 5
Db 17579 NITQLSINDCNRDAGKITLKLNSGSKSAFYTVKVIDTPGPONLAVEXRKSAPLV 17638
QY 6 ----- 5
Db 17639 WEPPIIDGAKVNYVIDRRESTRKAYANVSSKSTSEKVENLEGALIYFRVNAENEP 17698
QY 6 ----- 5
Db 17699 GVGVPVEIVDAKAAEPSPGKVTILTDVQSOTASILMENEKPEHDGSKVLGIYVEMQK 17758
QY 6 ----- 5
Db 17759 TEKMSIYASKVCMAVVTGLSSGQEQYFRKAYNEKSDPRLGVYLAKDITQPSLK 17818
QY 6 ----- 5
Db 17819 IAPNTYSIQAGEDLKIEIVTGNRPNISWVADGELKQTRVNEBETSTVLNIKEGN 17878
QY 6 ----- 5
Db 17879 KDDFGKITVTAINSAGTATENISVIYLEKPPGVPRDEVASDPVVISNPPAYTGCC 17938
QY 6 ----- 5
Db 17939 QISNTIYERKDDTTTWWVDSATVAATITLKVTKLGTIEQPIFAENRYGGSFALSDSKA 17998
QY 6 ----- 5
Db 17999 VIVQYPRKEPBPPTPATAISKDSMWIQWHEPVNNGGSPVIGYHLEKERNSLMTKVNT 18058

```

0Y 6 ----- 5
Db 18059 KPIJDIFKTTGDEGLEEFKVAENIVGIGKPSKVECFVANDPCDPPGPEAIYIT 18118
0Y 6 ----- 5
Db 18119 RNNYTLAKKKPAYDGGSKITGIVYEKKDLPDGRMKKASFTVLTETETVSGLVEDQRYEF 18178
0Y 6 ----- 5
Db 18179 RVIAHNAAGNFSEPSDSSGALTARDEIDAPNASIDPKYKDYIVYHAGETFYLEADIRGKP 18238
0Y 6 ----- 5
Db 18239 IPDVAWSKOGKELEETARMBIKSTIORTLVVDCIITDGGYILKLSNVGTRKSIPIIT 18298
0Y 6 ----- 5
Db 18299 VKYIDRPSPEGPLKYGVTAKECYLANNPLODGGANISHYIIEKRETSRLSTOVSTE 18358
0Y 6 ----- 5
Db 18359 VOALMYKVKILLPGNEYIFRWAVANKYIGEPLESGPYACNPYKPPGPSTPEVAITK 18418
0Y 6 ----- HSMEXR 11
Db 18419 DSANVTARPVDDGTEIEGYILEKROKEGVMTCKNKTITDILRLVYTGITGHSYEF 18478
0Y 12 ----- 11
Db 18479 VAAEHAAGVGESEPSVRYRACDALYPPGPPSNPKYTDTSSSVSLAKSKPIYDGGAPVK 18538
0Y 12 ----- 11
Db 18539 GYVVEKKAADWTTCTPTGLOGKQVYKLEKTEYNERICAINSEGVGEPAITPGS 18598
0Y 12 ----- 11
Db 18599 VVAQERIEPEIEADIRKVVYLRASATLRLFYITKGRPEPVKWKARAGIILDRAGIE 18658
0Y 12 ----- 11
Db 18659 VTSFTMLVDNTRFDSGRYNLTLENNSGSKAFVAVRILDSAPVNLIRVKKDSV 18718
0Y 12 ----- 11
Db 18719 TISWEPLIDGAKITNIVYEKRETKRAYATINNCTKTFRIENTLOEGSYFRVLAAS 18778
0Y 12 ----- 11
Db 18779 NEYIGLPAETTEPVKSEPLPGRYTLVDYTRNTATIKWEKPESDGSKITGYVEMQ 18838
0Y 12 ----- 11
Db 18839 TKGSEKSTCTQVTELEATISGLAGEYFVRVAVNEKGRSDPRLGVPIVARIIDIKP 18898
0Y 12 ----- 11
Db 18899 SVELPHTFNVKAREQKIDVPEGRPOATVNRKDGQLEKTRVAVSSKVTWLSLIR 18958
0Y 12 ----- 11
Db 18959 EASKEDVGYELCVNSAGSIYVITIVDRGPPGPRIHIDEVSCDSITISHNPEYDQ 19018
0Y 12 ----- 11
Db 19019 GCOISNTIVKETTSTTHIVSOAVARTSIKIVRLTGTSEYOFVCAENRYSKSSSES 19078
0Y 12 ----- 11
Db 19079 SAVVAEPPSPGPPGPKYVVAHAKSTMLVTQVPPVNDGSRVIGYHLEKRSILMSK 19138
0Y 12 ----- 11

Db 19139 ANKILLADQVAVSGLEGLMEYRVYAENIAGIKGCSKCEPVAPARDPCDPPGPEVTN 19198
0Y 12 ----- 11
Db 19199 ITRKSVLSKMSKPHYDGAKITGYIVERRELPGRWLKNYNTIOETFEVETELTEDQRY 19258
0Y 12 ----- 11
Db 19259 EFRVFAHNAADSVSESESTGPIIVKDDVPEPPVMDVYFRDYIVVAKGEVTLKINADIA 19318
0Y 12 ----- 11
Db 19319 RPLPVISMADGIEIEBARTETISTDNHTLTVKCOIRRDNGQVYVTLKNAGTFSVAV 19378
0Y 12 ----- 11
Db 19379 NCKVLDKPGPPAGPLEINGLTAEKCSLSWGRPOEDGADIDYVHRKRETSHLAWTICEG 19438
0Y 12 ----- 11
Db 19439 ELQMTSCVYKLLKGNIEYIFRYTVGNKYGVGEPLSVAILKALDPFTVPSPPTSLEITSV 19498
0Y 12 ----- 11
Db 19499 KESMTLQWSRPESDGSEISGYIIEERREKNSLRVAVNKKPYDLAKVSTGLREGEYEF 19558
0Y 12 ----- 11
Db 19559 RYTAENNAGLSSETSPLIRADPVFLPSPSKPIVDSGRITITIAMVKPLDFGAPI 19618
0Y 12 ----- 11
Db 19619 TGYTVBKSSDDIDAKTISQSLACTEYITISGLTGAIEYFRVKSVAKVAGASDPSDSDPO 19678
0Y 12 ----- 11
Db 19679 IAKEREPEPLFIDSEMKRTLYKAGASFTKVPFGRPPVAVMSKPDIDLRNAVYDI 19738
0Y 12 ----- 11
Db 19739 TDSRTSLTIEMANRNDSGYITTLIONVLSAASLTIVKVLDPGPHITVQDVKESAV 19798
0Y 12 ----- 11
Db 19799 LSWDPENDGAPVKNYHIEKREASKKAVSVTNONRSLSVTNLOEGAIYFRVSGEN 19858
0Y 12 ----- 11
Db 19859 EFGVGIAPATKEGVKITEKSPPEKLGVTJSKDSVSLTWLKEPHDGGRIHYVVALE 19918
0Y 12 ----- 11
Db 19919 KQKNNVCAVAKSTHVVVSGLRNSETFRVFAEHQGLSDPRELLPVILKEQLEPPE 19978
0Y 12 ----- 11
Db 19979 IDAKNEPSTVYVAGSNLKVDPISGRPLPVTLISROGVPLKATMRNTEITANENLIN 20038
0Y 12 ----- 11
Db 20039 LKESVADAGREYITANSSGTTKAFINIVLDRGPPGTPVYISDITEESVTLMEBPK 20098
0Y 12 ----- 11
Db 20099 YDGSQVNTYILLKRETSYAVTEVSATVARTMAKVMKLTGEEYOFALKAENRGISDH 20158
0Y 12 ----- 11
Db 20159 IDSACTVALLPYTPGPSTIPVNTVNTRESITVGHEPVSNGSAVGVYHLEKMDRNSIL 20218
0Y 12 ----- 11

Db 20219 WOKANKLVIPTTHFKVTVIISAGLIYEEFVVAENNAAGVKBSPSEPVLAIDACEPPRNR 20278
 QY 12 ----- 11
 Db 20279 ITDISKNSVLSKQOPAFDGSKITGYIVERNDLPDGRMTKASTNTNTEIOFTISGLTON 20338
 QY 12 ----- 11
 Db 20339 SOYERVFARNAVGSISNPSWVGPITCIDSYGGVIDLPLEYTEVVKRAGTSVKLRAG 20398
 QY 12 ----- 11
 Db 20399 ISGRAPPIEYKDDKELQNALVCVENTDLSILIKADRLNSGCEILKRNAMASAS 20458
 QY 12 ----- 11
 Db 20459 AFIRQIILDKGPPROGPIEFKIVTAETTLRMRPADOGAKITHIYVEKRETSRVVSM 20518
 QY 12 ----- 11
 Db 20519 VSEHLEECIITTKIKNGEYIFRVAVNKYIGEPLESDVAKNAFVTPGPGIPEVT 20578
 QY 12 ----- 11
 Db 20579 KITNSMTVNSRPIADGSDISGFIEKRDKSLGMEFKLKETIRDTROKVTGLTENS 20638
 QY 12 ----- 11
 Db 20639 YQYRCVAVNAGQGFSPSPFKAADPIDPPGPAKIRIADSTKSSITLGMSKPYVDG 20698
 QY 12 ----- 11
 Db 20699 SAVTGYVEIKQGEETVSTKGEVTEREVVSNLPGVNYFVRSVAVACQGEPIE 20758
 QY 12 ----- 11
 Db 20759 MNEPVQADILEAPIDIDVALRTSVAKAGEDVQVLIPEKGRPPYVTKKDEKNTGSD 20818
 QY 12 ----- 11
 Db 20819 ARSTENTDSSILITPQVTRNDGKIITLENGVEPKSSTVSKVLDPAQKQV 20878
 QY 12 ----- 11
 Db 20879 HVSNGVTIAMDPLIDGSPINIVIEKRDATKRTSVSHKCSSTPSFKLIDSEKTPF 20938
 QY 12 ----- 11
 Db 20939 FFRVLAENEIGICEFCTTEPVKAAEVPAPIROLSMADSTKSVILSWTAFDFDGGSVIT 20998
 QY 12 ----- 11
 Db 20999 EXVVRKMGQEQWWSHAGISKTCEIVSQLKQSVLEFRVAKNKGLSDPVTIGPITVK 21058
 QY 12 ----- 11
 Db 21059 ELITPEVDLSDIPCAQVTVRIGNNVHLEPKGKPKPSISWLDGLPLKESEFVRESKT 21118
 QY 12 ----- 11
 Db 21119 ENKITLSTKNNAKENGKTYTILDNAVCRIAPVITVILGPSKPKGPIRDEIKADSVI 21178
 QY 12 ----- 11
 Db 21179 ISMDVPEDNGGGEITCTSIKRETSQTNMKMYCSVARTFKPVLVDAEYQVRRAEN 21238
 QY 12 ----- 11
 Db 21239 RYGVSOPLVSSITVAKHOFRIJPGPKPIYVNTSDGKSLWDAVYDGGSEVTFHVEK 21298
 QY 12 ----- 11
 Db 21299 KERNSILMOKVNTSPISGREYRATGLVEGLDYQFNVYAEENAGLSPSDPSEKFTIIVASPV 21358

QY 12 -----TPD----- 14
 Db 21359 DPGPTDYIDVYREFITLKNPPLNDGSKIVGYSIEKQNGERNVRGCFDVSQCYTV 21418
 QY 15 ----- 14
 Db 21419 TGLSPDRYEFRILANNAVGTISPPSQSGIIMTRDENVPPIVEFGPEYFDGLIKSGES 21478
 QY 15 ----- 14
 Db 21479 LRKALVQGEHPVRYTWFDGYEIEKRNAMETVNLGSTISLEVNDATDHRGVTVEAKN 21538
 QY 15 ----- 14
 Db 21539 ASGSABRIKVKQDTPGKVGPPIRFTNITGEMTLMWDAPLNDGCAPIITHYIIEKRETS 21598
 QY 15 ----- 14
 Db 21599 RLAMALLIEDKCEAOSTAIKLINGEYQFVSAVNFVGRPLDSDPVVAQIQYTVDPAP 21658
 QY 15 ----- 14
 Db 21659 GIPESNITGNSITTLWARPESDGSSEIQYTLERREKSTRWKVYISKRPISERFKVT 21718
 QY 15 -----INP----- 17
 Db 21719 GLTEGNEYEHVMAENAGVGPASGISRLIKGRFPVNPGPPTVVKVDTSKTVLSLMS 21778
 QY 18 ----- 17
 Db 21779 KPVFDGMEIIGYIIECKCTDLGDMHKVNAECVKTTRYTVTDLQAGEEYKRVASLINGAG 21838
 QY 18 ----- 17
 Db 21839 KQDSCEVGTIKAVDRLTAPELDIDANFKQTHVVRGASIRLEIACQGPPTVAVMSKD 21898
 QY 18 ----- 17
 Db 21899 SNLSINADIHTTDSFTLVECNCRNDAGKYTLVENNSGSKSTPVTVLDPGPGPI 21958
 QY 18 ----- 17
 Db 21959 TFRDVTGSAITLMDAPLIDGCAIRIHVYVEKREASRSMQVISEKCTQIFRVNDLAE 22018
 QY 18 ----- 17
 Db 22019 VPIYFVSAVNEGYGEPEYEMPEPIVATEQAPAPRRLDVVDKSSAVLAMLKPDHDGGS 22078
 QY 18 ----- 17
 Db 22079 RITGYLLENKQSGDLMVEAGHTKQITVVERLVENTEYEFVKAKNDAGSEPREAPSS 22138
 QY 18 ----- 17
 Db 22139 VTIKROIEFTADLTGITNOLITCKAGSPFTIDVPISGRPAKVTYKLEEMRLKETDYS 22198
 QY 18 ----- 17
 Db 22199 IITTKDRTITLVKSMRGDGRFLENTAGKFTSVTVVIGRGPYTGPLEVSSVSA 22258
 QY 18 ----- 17
 Db 22259 ESCVLWGBEPKDGGEITNTYIEKRESGTTANOLVNSSVKRTOIKVTLTRIMEYSFRV 22318
 QY 18 ----- 17
 Db 22319 SSENRRGVSKPLPSAPLIAEHFVPPSAPTRPEVYHVSANANSIRNEEPIHDGSKTIIGY 22378
 QY 18 ----- 17
 Db 22379 WVEKKENTITLWKENKVPCLBENKYVTGLVEGLEQFRTYALNAGVSKASEASRPIMA 22438


```

OY 18 ----- 17
Db 22439 ONPDAQRPEVDTVNSTVSLWSAPAYDGSKVGYIIRKPVSEYGDGRWLKCNVTI 22498
OY 18 ----- 17
Db 22439 VSDNETYALSESDTYEFRLAKNAGVJSKSESTGPTWCDEYAPPAKELARLHGD 22558
OY 18 ----- 17
Db 22559 LVTIRAGSDLVDAVAGKPEPKIIVTKGRKELDCENVSIOYTKKATINIKRCDSDS 22618
OY 18 ----- 17
Db 22619 GKTYLVKNASGTAVSVMKVLDSPGCKLTVSRVQCKTLAWSLPQDGAETHX 22678
OY 18 ----- 17
Db 22679 IVERRETSRLMWIYEGCPTLSTVYVRLIKNNETIFRRAVNKYGGVPESEPTIARN 22738
OY 18 ----- 17
Db 22739 SFTLPSPGIDEYGTGKEHIIIOWKPESDGNEISNLYDKRKSLSMTIRVNDYV 22798
OY 18 ----- 17
Db 22799 YDRLAKVTSIMEGCDYOFRTAVANAGNSEPERSNFI SCHEPSTPGPPSAPRVDTTK 22858
OY 18 ----- 17
Db 22859 HSI SLAWTKPKYDGTDIYGVYVLEMOEKDIDQWYRHTNATIRNTEYVBDLKKGOKYSF 22918
OY 18 ----- 17
Db 22919 RVAANVYKGMSEYSESIAREPVERIEIPDLADLKTVTIRAGASLRMWSGRPP 22978
OY 18 ----- 17
Db 22979 PVITWKSOGIDLASRALIIDTESISLIYDKVNRDAGKTTIENOSGKKSATVLWVY 23038
OY 18 ----- 17
Db 23039 DTPGPCSVKKEYSRDSVTITWEIPTIDGAPINNIYKREARNAKFTVTKGSKTL 23098
OY 18 ----- 17
Db 23099 YRISGLVEGTMHFRVLPENLYGIGECETSDAVLYSEVPLVPAKLEVVDVTKSTVLAW 23158
OY 18 ----- 17
Db 23159 EKPLIDGGSRLTGYVLEACKAGTERMAKVYTLAPTYLEHTVTSINSGEOTLFRIRADNEK 23218
OY 18 ----- 17
Db 23219 GVSEPRETVAVVODLEKPLTIDLSMPQKTIHVPAGRVLEVIPIAGRPAPASWFEA 23278
OY 18 ----- 17
Db 23279 GSKURESERVETHTKVAALLTIRETTIRDTGTYLELKNVTGTSBTIKVILLDRKPP 23338
OY 18 ----- 17
Db 23339 TGPPIKIDEIATSIITISWEPBELDGAPLSGYVVEQDAHRGMLVSESVTSTKFTIR 23398
OY 18 ----- 17
Db 23399 LTBGNETVFRVAATNNRGISYLSQSEVIECRSIRIPGPETILOIDVSRDGLTWYPP 23458
OY 18 ----- 17
Db 23459 EDDGSGOVTVIYERKEVRADRVRNKKVYVYTWIRYRSTIGLREYERHVTAINARSGS 23518
OY 18 ----- 17

```

```

Db 23519 KPSNRSPKIYAMPPIAPPGKQPNRVTIDITRTSVLSLMSVPEDEGSKVTGILIMQKVD 23578
OY 18 ----- 17
Db 23579 QHEWTKCNTTPTKIREYTLTHLPQGAERYRRLACNANGRGEPAEYGVTKVTEMLEYPD 23638
OY 18 ----- 17
Db 23639 YELDERIOESIFVRQGVIRLTIPIKGRPEPICKTREGODISKRAMIATSETHLVLK 23698
OY 18 ----- 17
Db 23699 EADRGDSGYDVLLENKCKGKAYIKVRVIGSPNPEGLPYNDIOVRSVRSNRPAD 23758
OY 18 ----- 17
Db 23759 GGADILGYTLERREVPRAWYTLDSHRGTSLVYKGLKENVEYHFRVSAENQFGLSKPLK 23818
OY 18 ----- 17
Db 23819 SEEPVTPKPIANPPEPPSNPEVLDVTKSSVLSMSRPRDGSRTVGYIERKETSTDK 23878
OY 23 ----- 22
Db 23879 VVRHNKTQITTYVTVGLVPAEYQFRIIAQNDVGLSETPASEPYCKDPDKPSQPG 23938
OY 23 ----- 22
Db 23939 ELEIISKDSVTLQWEKEDCGKEILGYWEYRQSGDSANKSKNERIKDKOFTIGL 23998
OY 23 ----- 22
Db 23999 LEATEYEFVFAENETGLSRPRTAMSITKLTSGEAPGIREMDYTTKLGEAOLSCQ 24058
OY 23 ----- 22
Db 24059 IVGRPLDIRMYRFGKELIOSRYKMSDGRTHLLVMEQEDGCVYTCIATNEVEYE 24118
OY 23 ----- 22
Db 24119 TSKLLQATPOPHGCIPIKEKYAGVSTLRLAWYIGRPVAMTWFGOKLIQNSBN 24178
OY 23 ----- 22
Db 24179 TIENETHYHLVKNVQRKTHAGKYKVOLSNFQVDAIIDVEIODKDPGPIVEAL 24238
OY 23 ----- 22
Db 24239 LKNSAVISWRPADDGSMITNYVEKCEAKGAEOLVSSAISVTCRIVNLTEMAGY 24298
OY 23 ----- 22
Db 24299 FRVSAONTGIDPLEVSSVYIISPFKPGAPGPTITANTKDCSVVAMKPPASDGAK 24358
OY 23 ----- 22
Db 24359 INNYLEKREKONRWISVTEIREFVSVKNLIEGLEEFVACENILGSESEISE 24418
OY 23 ----- 22
Db 24419 PTPPSDVPICAPPHKEELRNINVRYSNAILVCRVTHPKPRIYKRYGKEI IADGLKY 24478
OY 23 ----- 22
Db 24479 RLOEFGYHOLIASVTDDBATVYQVATNGSGSVSTASLEAVVPKHLPTLBSMG 24538
OY 23 ----- 22
Db 24539 AVHALGEVSLIKIDPSKPDVITWOGODLIDNNGHYQVIVTMSFTSLVPMGVERKD 24598
OY 23 ----- 25

```

Db 24599 AGFYVCAKNRFGIDOKTVELDADVPDPGRGVFVSASRDSVNLTWTEPASDGSKITN 24658
 QY 26 ----- 25
 Db 24659 YIVKCAITTAERMLRVGQARETRTYVNLFGKTSYQRYVAENKFGLSKSEBPSEPTITK 24718
 QY 26 ----- 25
 Db 24719 EDKTRAMNYDEVDREVSMTKASHSSTKELYEKYMAIEDLGREGFIVHCVETSSKK 24778
 QY 26 ----- 25
 Db 24779 TYMAFVVKGDVLYNKEISILNIAHRNLIHLHSFESMEELVMTFFISGLDIPER 24838
 QY 26 ----- 25
 Db 24839 INTSAFELNEREIVSYVQCEALQFLSHNIGHFDIREPNIYQTRSSYIKIIEFGQA 24898
 QY 26 ----- 25
 Db 24899 KOLKPGDNFRLFTAPEYAEVHQHDVSTATDKMSISGLTVYLLSGINFLAETNQOI 24958
 QY 26 ----- 25
 Db 24959 IENINNAEYTDDEAFKEISIEAMDEVDRLLVKEERKSHPTASEALQHPWIKOKITERVSTK 25018
 QY 26 ----- 25
 Db 25019 VIRTAKHRYVHTLKKDLNMYVSARISCGAIRSOKGSAVAKVASISIGPVSOIM 25078
 QY 26 ----- 25
 Db 25079 HAVGEGGHVYVCKIENYDOSTQVWYFVGOLENSHKEIYEDGVAIIVYKDITKID 25138
 QY 26 ----- 25
 Db 25139 DGTTCYKAVNDYGDSSYAEIYKGVREYDYICRTYKTKIKRTIDIRKLLERPEFTLP 25198
 QY 26 ----- 25
 Db 25199 LYKNTAVYGENVBSVTLTVHPEPHVTWYKSGQKIKPGDNKKYTFESDKGLYQLTINSV 25258
 QY 26 ----- 25
 Db 25259 TTDDAEYTVARNKYGEDSCAKLTVLHPPTDSTLRPMFKRLLANAEQEQSVCFE 25318
 QY 26 ----- 25
 Db 25319 IYVSGIIPPTLKNKEDGQPLSIGPNIETIHBEGLDYALHINDTLPEDIGYRVATINTAG 25378
 QY 26 ----- 25
 Db 25379 STSCQAHQVERLYRKQOFKSEHEHRYVQKIDKTLMAAILSGTESVPLTQVAKKAL 25438
 QY 26 ----- 25
 Db 25439 KRAAVLYKPAVSTKTVKGFLEIEKKEERKLKMPDYVPEPRKYKQTTIEODQIKOFV 25498
 QY 26 ----- 25
 Db 25499 PMSDKKTKYKIDQYEMPKLDRVQKPKRIRLSRMEQFTYMPLPRTDQYRPKWRIPK 25558
 QY 26 ----- 25
 Db 25559 LQODDLEIVRPARRTPSPDYFYRPRRSIGDISDETLPLIDDYLA MKTEERLRL 25618
 QY 26 ----- 25
 Db 25619 EEELEIGFSAPPSRSPPHFELSLRYSSPOAHVVEETRNKFRSYTHIPTKABASTSY 25678
 QY 26 ----- 25
 Db 25679 AELREHQAQAVRQKORIRIAREDEELLPVTTTQHLSTYKSELDFMSKEEKSRRKS 25738

QY 26 ----- RPV----- 28
 Db 25739 RQREVEITEIEBEVEYKSHAORESSSASRLRRBSLPTYLEMRPVSELIRSRPQ 25798
 QY 29 ----- 28
 Db 25799 PAEEYDDTERGSPTEPRTPSPSPVSSERLSRERSARDIFSRYESMAKALTKOXT 25858
 QY 29 ----- 28
 Db 25859 SEKRYEVLQOPPTLDHAPRTILMRSHRVPCGQNTRFILNVQSKPLAEVKNVHGVELQ 25918
 QY 29 ----- 28
 Db 25919 ESSKIHVTNTSGVLTLEILDCHTDGGYRAVCTNYKGBASDYLDTVGDYTYASQR 25978
 QY 29 ----- 28
 Db 25979 RDEEVPBSVPELTREAVAVPSFKYTSMEASSSVREYKSOQTEPRBSLSYEHSAAE 26038
 QY 29 ----- 28
 Db 26039 MKSNALEKSEKSEKSTTRIKITLARIILTKPRSMVYEGESARSCDIDGEPVPTVYL 26098
 QY 29 ----- 28
 Db 26099 RKGVLSTSAHQVTTTKYKSTFEISSVQASDEGNYSVVENSECKQAEFTLITOKANV 26158
 QY 29 ----- 28
 Db 26159 TEKAVTSPPRVKSPEPRVKSPEAVKSPKRYKSPBSPKPAVSFTEKPPPIEKVOHLPS 26218
 QY 29 ----- 28
 Db 26219 APPKITQFLAKASKELAKLTCVRESSYLAKREVWYKDGKLNKENGHPQFHSADGYE 26278
 QY 29 ----- 28
 Db 26279 LKINLTESDQGEYVCEISGEGTSKTNLQFMQAFKSIHERVSKISETKKSDQKTTST 26338
 QY 29 ----- 28
 Db 26339 VTKTEPKAPEPISSKYIVYTGLODITVSSDSVAKAVATGPRPTAINTKDKALITOG 26398
 QY 29 ----- 28
 Db 26399 GKYLSEDKGFLEIHKTDTSQGLYCTVKNSAGSVSSCKLTITKAKDPEAQVSTQ 26458
 QY 29 ----- 28
 Db 26459 KTSEITPQKAVVQEBISORALNSEETKSEAKSOEKLAKEBASVYLISEVYKSAATS 26518
 QY 29 ----- 28
 Db 26519 LEKSTVHEELTKTSQASEEYRTHAEIKAFSTQMSINQGLVYKANTAGATDVKWVLANCV 26578
 QY 29 ----- 28
 Db 26579 ELTNSEERYGVSGSDQTLTIKQASHRDEGLIINCISKTEGIYKQYDLTILSKELSDABA 26638
 QY 29 ----- 28
 Db 26639 FISQPSQNTINEQNTLFTREISGEBSPLEIEMFKNNLPISISSNVISISGRNVYSLEIN 26698
 QY 29 ----- 28
 Db 26699 ASVSDSGYTIKAKNFRGQCSATASLMLVLPVPEPSHEVYLRTSGDTSLOGSSSSQVOM 26758
 QY 29 ----- 28
 Db 26759 SASQOASFSFSSSSASNTKEMKFAQMSAQMSMOESFEVMSSSSFMGISNTOLES 26818

QY 29 -----G 29
 Db 26819 TSKRLKAGIRIPRIKIALPSIDISEGKVLTVACAFTEPTPEVTWSCGRIHQEOG 26878
 QY 30 RF 31
 Db 26879 RF 26880

RESULT 5
 T42629
 tenascin-X - bovine
 N:Alternate names: flexillin
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T42629
 R:Eleftheriou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C.
 J. Biol. Chem. 272, 22866-22874, 1997
 A:Title: Characterization of the bovine tenascin-X
 A:Reference number: 222180; MUID:97426436; PMID:9278449
 A:Accession: T42629
 A>Status: Preliminary; translated from GB/EMBL/DDAJ
 A:Molecule type: mRNA
 A:Residues: 1-4135 <ELEX>
 A:Cross-references: EMBL:Y11915; NID:92462978; PIDN:CA72671.1; PID:92462979
 C:Genetics:
 A:Gene: TN-X
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin tyh
 C:Keywords: extracellular matrix; glycoprotein; heptad repeat

Query Match 41.0%; Score 64; DB 2; Length 4135;
 Best Local Similarity 0.4%; Pred. No. 2.4e+03;
 Matches 17; Conservative 5; Mismatches 8; Indels 4048; Gaps 4;

QY 1 SRKHX----- 6
 Db 56 SRLHEFTVEGGEKQVFTHRINLPPSAGCGCPOTEPPVPASEVQALRVRLLEELVKG 115
 QY 7 ----- 6
 Db 116 LKEDCTGGCCPPAAQAGTGQTDIRSLSLHGVFDLSRCACSCPEWGGPTCSPEGAGP 175
 QY 7 ----- 6
 Db 176 PSSPSAPSGCPDDCNDQGRVCYGRKCYCFPGYTPSCSNPSCPGDCHGRGRCVQGVCR 235
 QY 7 ----- 6
 Db 236 SGFSGDDCVSRSCPRGCSQGRGCEDEGRVCNPGYSGEDGVSRCPDSCQNGRCNGRCV 295
 QY 7 ----- 6
 Db 296 CNPPTGDDCGVSRCPGCSQKGRCEDEGRVCNPGYTGDDCGSRTCPWDCBGRGCVDR 355
 QY 7 ----- 6
 Db 356 CYCMFAGYAGDCSTRCPDRCRPRGCEDEGCICDPIGSDGCVSRCPDSCNORGRCE 415
 QY 7 ----- 6
 Db 416 GRVCWPGYSGPDGAGAPRDCRGRRCENGVCYCHAGYSGEDGVSRCPDCCRRGRK 475
 QY 7 ----- 6
 Db 476 ESRKCYWPGYTGRCDCSTRACPDCCKGRGRCVDRVCNPGFAGEDCGSRCPDGRGK 535
 QY 7 ----- 6
 Db 536 RCGGVSCDVYGEDEDCGRKSRCPRGQGRGQGLEGRVCYDDGYEGEDGVSRCPDCAQ 595
 QY 7 ----- 6
 Db 596 KGVCQDCVTCWNGFAGEDGRLVCPNCHRRGRCENGRCVDSGYTGPSCATRTCPAD 655

QY 7 ----- 6
 Db 656 RGRRCVQGVCHVCHVYSGEDCGEPPASACPGCGPRELCSAGCCVCEGFRPDDAI 715
 QY 7 ----- 6
 Db 716 QMCPDGRGRKRCRSCVCQDGYAGEDCEEPALIEGRMHLLEETVTEMTAPGNV 775
 QY 7 ----- 6
 Db 776 DAVEIQPIPTBESAPPTARVPSSASAYDKGLAFGEYQVTRALGCTNGPPASKTI 835
 QY 7 ----- 6
 Db 836 TTMIDGRQDLRVAVVPTTLELNLRLPQADNDRFVTVSYSAQNGRVLEVPSEADGTLIT 895
 QY 7 ----- 6
 Db 896 GLMPGVYVTVLBERGRAVSYPASIRANTGSSLGLGATDEPPSPGPTQGAQAPVL 955
 QY 7 ----- 6
 Db 956 QQRPELAEIRVLGKDKTGLRVNMTAQPDTTHRQLRLVEEGGANHEELPGDVQAL 1015
 QY 7 ----- 6
 Db 1016 VPSPPESPVELSLRGIPPSGESAFLIYQIMDKCEKPKPLAPRLGLVTVDITSD 1075
 QY 7 ----- 6
 Db 1076 SLLMTVPGEERDFSVIQKDRDRQVVPVGPQKSAISNLDVGRKKEVLYGLVGRK 1135
 QY 7 ----- 6
 Db 1136 RHGPIVAEKILISQDDPSVPTPRIGNIMVTPPTDLSLSTWVEGQDFSNQYRRA 1195
 QY 7 ----- 6
 Db 1196 GRQVYVVEGPPRSVLIPLDPDHVRFLLFGIANKNRHPLADGTAPEKKEPRRPE 1255
 QY 7 ----- 6
 Db 1256 PPRPLIGELVAGATASLSLSTWVAQSFDSFVYQKDAQGRQAVVTDENEVAIP 1315
 QY 7 ----- 6
 Db 1316 SLEPDKKYKMLXGLGRQRVGVSVAITAPQVLDETPSNTEMEETPSPTSTAPE 1375
 QY 7 ----- 6
 Db 1376 SPEKPLIGELAVTGSSPDSLISLWTLPGHFDSTFYQYRDGDQPRVNRVGDDEGVITIS 1435
 QY 7 ----- 6
 Db 1436 GLEPDHKYKMLYGFHDROKRGPMKSVIGVTAEEETPSPTMEETPSPTVEETPSPMER 1495
 QY 7 ----- 6
 Db 1496 STEAPEAPEPLIGELVITGSSPDSLISLWTVQGHFDSTFYQYGRGPOVYKRGRET 1555
 QY 7 ----- 6
 Db 1556 EVTIGLEPGRKYNLXGLHGRQSGVSTVGVTAPQAEETPATEPPKPRIGELVIT 1615
 QY 7 ----- 6
 Db 1616 DVTNSVGLERTVSEGQDSEFVYQKDRDQSHVVPVADIDREATVSGLEPERKRYRANVY 1675
 QY 7 ----- 6
 Db 1676 GLHGQRVGPLSVALLTAPVPPDPVTEPPVEPRIGELVITVTPDSVGLSTWVAGEERDS 1735

```

QY 7 -----SMEEXT----- 12
Db 1736 FLVDYKDRDGOPOVVPATDQREVTLPGLEPSRKYKFLFGIDGKRRTSVSEAKTVGR 1795
QY 13 ----- 12
Db 1796 GDASRGAPRLGELMTDPTPDSLRLSWVPBGHPDSFVVQFKDRDGRVRSVEGHERSV 1855
QY 13 ----- 12
Db 1856 TISPLDGRKRYRLVYGLGKRKRKPLTTEGTTETRRVADAGTKRPSKPLGELQVYG 1915
QY 13 ----- 12
Db 1916 VTSNSVGLSTVPRGRDSEVYQTRDNDGQPOVVPVEGSRREVSGLDPAKRLKLLYG 1975
QY 13 ----- 12
Db 1976 LSRDKRVPTSAIVTEPAPREIKAPATFSPASRPLGELYLEAPHSRLSWTAT 2035
QY 13 ----- 12
Db 2036 EGEDSEFVOTDENGOLOEVNVGGQHDITISDLESDHRYLSLYGHDGQVGAHIE 2095
QY 13 ----- 12
Db 2096 AMTAPREDEPSESLSSTQTPSTAYPEPHIKRGLGELAVTDTPDLSLSWTVPEGQFD 2155
QY 13 ----- 12
Db 2156 HFLIQKNGDQPKVYRPGDEDEVTISGLEPDHKYKMLYGFHNRORMPVSVIGVTTA 2215
QY 13 ----- 12
Db 2216 EDEPSPTEMEETPSPTEMEETPSPTEPNTVEPPPEPPELELTYTGSSPDSLSTSWY 2275
QY 13 ----- 12
Db 2276 PQGHFDSFTVQYKNGDQPKVYRPGHDKGTISGLEPDHKYKMLYGFHDROVGPVS 2335
QY 13 ----- 12
Db 2336 IGVTAEEETPSPTEVEETPSPTEPSTEAPPEPPEVLGELMTVGTSSPDSLSTSWTVPOG 2395
QY 13 ----- 12
Db 2396 RFDSEFTVQYKGRDGPQVYRVGGEETEVTBGLERGHYKMLYGLHGRVGPSTVAMT 2455
QY 13 ----- 12
Db 2456 AREEPASPLKPOLGELVTVDATPDSLSTSWTVPEGQFDHFLYQKNGDQPKAVRVP 2515
QY 13 ----- 12
Db 2516 GDEGVTISGLEPDHKYKMLYGFHDROVGPSTVGLTVEKQDEMTAPATDLPATAPE 2575
QY 13 ----- 12
Db 2576 PPIKPRGELVVDATPDSLSTSWTVPEGQFDHFLYQKNGDQPKAVRVP 2635
QY 13 ----- 12
Db 2636 GLEPDHKYKMLYGFHDROVGPVSVIGVTAEEETPSPTEMEOTPSPTEVEETPGTEV 2695
QY 13 ----- 12
Db 2696 EETSPPTPSTEAPPEPPELGLGELTVTGSSPDSLSTSWTVPOGHPSFTIQHKGRCPO 2755
QY 13 ----- 12
Db 2756 VVRVGGEETEVTIGGLEPGRYKMLYGLSHGQVGPVSTVGTDPQEVVEETPSPTEPS 2815
QY 13 ----- 12

Db 2816 TEAPPEPPELGLDLYTVGSSPDSLSTSWTVPOGHFDSFTIQYKGRDQPOVVPGEETE 2875
QY 13 ----- 12
Db 2876 VTIGLEPERRKRYKMLYGLHGRVGPVSVIGVTAEDYDAMTQTPSTSPTEPPTKPLIG 2935
QY 13 ----- 12
Db 2936 EYVTVDATPDSLSTSWTIPEGQFDHFLYQKNGDQPKAVRVP 2995
QY 13 ----- 12
Db 2996 KMLYGFHDHQRVGPVSVIGVTAEEETPSPTEMEETPSPTEPSTEAPPEP 3055
QY 13 ----- 12
Db 3056 EEPPLGELTVTGSSPDSLSTSWTVPOGHFDSFTVQYKGRDGPQVVPGEETEVTVGLE 3115
QY 13 ----- 12
Db 3116 PGKRYKMLYGLHGRVGPASTVGTASTLTERPLARGLGELAVAVTSDPAHLSATVE 3175
QY 13 ----- 12
Db 3176 QGPDSFVQYKNGDQPOVVPAGDLREVTVSSIAPCRKYKFLFGLEDRGRHGPVSAD 3235
QY 13 ----- 18
Db 3236 ANTLPOTKAPRIGELVTVDTPGSGVGLSWTVEEEDFSFVQYKDRDGPVVPVADQ 3295
QY 19 ----- 18
Db 3296 REYTVPGLEPNKRYKFLYGLYGRKRLSPISAGSTAPLEKERQPPRLGELVTVDTPN 3355
QY 19 ----- 18
Db 3356 SLRLSWTVAGRRPDSFVVQYRGTDGPRMVPVADQRETVBGLERGRYKFLYGLIG 3415
QY 19 ----- 18
Db 3416 QRLGPASVIGMTAPEDDTAPAHMAATEAPKPEGRIGVLAANDVSPDSLSTSWVOP 3475
QY 19 ----- 18
Db 3476 FDSFVVQYODTDGQPOLAVGDQNKVLYSGLPSTSEFFLYGLHGRRLGPVABGTT 3535
QY 19 ----- 18
Db 3536 GPVPAGOTPEGPRLSHLSVTDVTTSSLRLNWEAPPEAFDSFLRFVPSPTLEPOLR 3595
QY 19 ----- 18
Db 3596 PLLQRELTVPGRRSAVLADLHPGILYTLTYGLNGPHKADSIOGTARTLSVLESPPDL 3655
QY 19 ----- 18
Db 3656 QFSEIRTSARVSWPTSPRVDFKVSQOLADGGEPOSVQVGTQKLEGLIFGAQYEV 3715
QY 19 ----- 18
Db 3716 VSVKGFEESEPLGFLTVVPDGPYHLALNLTDESALLHKMPPOOTPVDTYVYKVTANGA 3775
QY 19 ----- 18
Db 3776 PSLQASAPGSANDYPLQGLVTHNTATILRLGRPNFTSPASITTTGLEANPOLAEKV 3835
QY 19 ----- 18
Db 3836 TPRTALLTWTAPVSPTYGLSENPBGOTOELLPGVTSQHLNGLPSPSTYSIWLAM 3895
QY 19 ----- 18

```

Db 3896 MGDFTPPVSTSTTGGRLRPFPPDCEEMONGVSTRTTTLPLNGRERPLNFCOMET 3955
 QY 19 ----- 18
 Db 3956 DGGMLVFORRMDGKTDFMRDMDYAHGFNGISGEFWLGNALSLTKAGDYSLRVDLRA 4015
 QY 19 ----- 18
 Db 4016 GDEAVPAQYDSFOYDADETRYRLHLEGYHGTAGDSMSYHSGSVFASARDPNNLLISCAY 4075
 QY 19 ----- WYXXRGIT----- RPYGR 30
 Db 4076 SYRGAATYRNCNRYANLNGLYGSTVDHGVSNRYTKGDFSVPTFKMLRPRSTRFLGR 4133
 ||:||
 RESULT 6
 S55505
 fatty-acid synthase (EC 2.3.1.85) - Brevibacterium ammoniagenes
 C:Species: Brevibacterium ammoniagenes
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 05-May-2000
 C:Accession: S55505
 R:Stable, H.P.; Wagner, C.; Andreou, I.; Schweizer, E.
 Submitted to the EMBL Data Library, June 1995
 A:Reference number: S55505
 A:Accession: S55505
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3063 <STD>
 A:Cross-references: EMBL:X87822; NID:9861017; PID:CAA61087.1; PID:9861018
 C:Superfamily: Brevibacterium ammoniagenes fatty-acid synthase
 C:Keywords: acyltransferase; coenzyme A
 Query Match 40.4%; Score 63; DB 2; Length 3063;
 Best Local Similarity 0.7%; Pred. No. 1.7e+03;
 Matches 19; Conservative 2; Mismatches 10; Indels 2559; Gaps 5:
 QY 1 SRX----- 4
 ||:|
 Db 163 SRHMLSVKRPVREMEVEYTAGDAIAVNGRVHAFALSTPEDLANTESNLQAASYNDD 222
 QY 5 ----- 4
 Db 223 ALERRIGSEINPVVDYLAVALPFHHSADDAADLVDAVACGGDAELARELADSLIV 282
 QY 5 ----- 4
 Db 283 QPHSWETVAGLNSTYLLSLDRGLSLITPLIAGTGKVVPAATPAERDNLATPTLPT 342
 QY 5 ----- 4
 Db 343 AVNYEKFSPKLISLPNGKSTTGTRESEMGSPILLGMPPTTMDPGIVAAAANGYWE 402
 QY 5 ----- 4
 Db 403 MAGGGYSDAFTINKGMWELLPEGRTAAPNMFDRYIMNLPQVTRICSKARANGAA 462
 QY 5 ----- 4
 Db 463 FTGVTICAGIPELDEAKELLDOULTSDGFPYISFKPGTKQIDCAVIAADNPTHVYIQT 522
 QY 5 ----- 4
 Db 523 EDHAGHSHWDLDEMLATYACAREHONLATYGGGHSRDRASEYLTGTWSTYGLP 582
 QY 5 ----- 4
 Db 583 IMFVDGVLCTVAMATKEATANDVKOLLVDFPGISPTNGGNGVGLADGGVSSQSHL 642
 QY 5 ----- 4
 Db 643 LADLHEINSAKASRATISPIEYDERDEITAAIDRTSKRYGDLSEMYEDVAVNF 702
 QY 5 ----- 4

Db 703 AERAYPMDPTMHRHDFHDLORVEARLNDADGDIETLPTLIDSEHAEVAKLAAVP 762
 QY 5 ----- 4
 Db 763 NATTRVNT RDEAMPETLIRKHVKPMPWTTAIDGLKEWFANDTIQAODPRYDAGVNI 822
 QY 5 ----- 4
 Db 823 IPGPVAGITKKNEPVANLGRFEDATTAALNDAGVAVELISRLASAKNAEFTLRNP 882
 QY 5 ----- 4
 Db 883 TIIMHGLIANPAYELPEEAFDIDDDGGAIRINDSYRONLPREQRPYKHYDIPVA 942
 QY 5 ----- 4
 Db 943 LSEAVATGASPVVDARLPKAVFDLAGVAGYSISEGDKITTELPRVIEGVSSENPYG 1002
 QY 5 ----- 4
 Db 1003 LVETSTLPLSTLIATATVTAALGTANAGTPDALVGCWPAITATLGTGLTEEGEPA 1062
 QY 5 ----- 4
 Db 1063 GUDFPIESGLAHVLDHYVDVRFLEHLAKGEGEGRIDVTSKASIASSHGRIVTV 1122
 QY 5 ----- 4
 Db 1123 EELMDAATQEVVAMOMQFALIRGRATGTSVPVSPWGGKSODKIETTPRSFVDAIV 1182
 QY 5 ----- 4
 Db 1183 TAPSDTTPALYSGDYNPITHTNARLVHLDAPLVHGMLSATQOHLAHHGVYGFY 1242
 QY 5 ----- 4
 Db 1243 SMYGVOINDEVEITVERVKGKIHAEVYCRIDGEVVSQGLMAQPRYAVYPGGCI 1302
 QY 5 ----- 4
 Db 1303 QAEGRGDRDASAAAREVWRADRTTRTAGSTROIIDNPTELVVGRKEVHPNGVL 1362
 QY 5 ----- 8
 ||:|
 Db 1363 HLTFYVALAVYAAQTERLEREDALGTNSKYAGHSLGETTALASLANIFDLAVIDY 1422
 QY 9 ----- 8
 Db 1423 YRGSAMGTVERDENGNSNYGMALRPMIGVPADQVEAYIAQTABRTGELEIVYNI 1482
 QY 9 ----- 8
 Db 1483 AGQOYSIAGTKAGLALKKKAHSVADRAYVTPGIDVPFHSOVLADGVAPAEKIDELP 1542
 QY 9 ----- 8
 Db 1543 ELLDDALVGRVPLVALPPELTQEFVDKYPLAPSGKIDNLKVEDDEQAPRLMTE 1602
 QY 9 ----- 8
 Db 1603 LLSWQPASPVWIEITOULLFEVDQIIEVGLASSPTLTLNLRSMIDLAGVDLPVFNVERD 1662
 QY 9 ----- 8
 Db 1663 ODQVVLQDQVAPASAFVYEGEATSTTAASETPGESAAAAADNTQAIPSAEQVTAEP 1722
 QY 9 ----- 8
 Db 1723 APSAPAGSTRADAPDLPTAAEALIVLFAFQNKIRDDQINDSDYBELNGVSSRRNQ 1782
 QY 9 ----- 13
 ||:|

```

Db 1783 ILMDMSAHNVPAIDGADADAVATLRRVTAAPGSPGTVLSEAITARLQLTGAGV 1842
QY 14 ----- 13
Db 1843 KPAVTSRTGTGWLPMMAHVAEAILLGSREDSVGGSLSTVPSAASKADVALVD 1902
QY 14 ----- 13
Db 1903 AAYQAVAAAGTSVSHGAASGAGGGVVDASALDAVADIVTGENGYLATTAARQVLAQGL 1962
QY 14 ----- 13
Db 1963 VEEAPEPEPDNTLFAVVEAELSGMEKVTVPSPDAKRAVLFDRRASAREDLARVALGE 2022
QY 14 ----- 13
Db 2023 IDLPVRFPGTGETIAQAKEMAMENASTGAKAKTAETLTAIAAAARELDGEFAGD 2082
QY 14 ----- 13
Db 2083 VALVTGAAPGSIAVALVERLLLEGATVITMASRVSSRKEFAKLYAARAIPGAALWYV 2142
QY 14 ----- 13
Db 2143 ANLRSYRDVALLDWIGNEQASVGENEVIKPKALPPLIAPPAAVSQSVADAGPQAE 2202
QY 14 ----- 13
Db 2203 NQRTLLMASVERTIAGLSNAQGVQTRCHITVLPSPNRGMFGDDAIGVRAALDAIILA 2262
QY 14 ----- 13
Db 2263 KMSAEMPGEGVTLQAQKIGWVSGTSLMGNDVLIIPAAEAGIHWMPDEISSQLISIAS 2322
QY 14 ----- 13
Db 2323 EESRAKAAEAPLELDLTGGLGSSNISISELAQAREDAEQAASGDNAAPAEPAATIP 2382
QY 14 ----- 13
Db 2383 ALPMTREVELPALPEGEVDVTTDLDMVYIAGVGEVSSMSGRTFEEAEYGLQRGAV 2442
QY 14 ----- 24
Db 2443 DLTAAGVLELAAWMTGLISMSNDPRAPWYDEGEVDEADITYAFRDEYVARSGIRTLTK 2502
QY 25 ----- 24
Db 2503 YNNVDOGSIDLTSVFLDRDIVFTVPTBOEALDIEADPSFTKLAEVDGENEVTRLKATA 2562
QY 25 ----- 24
Db 2563 RVPKRALTRIVAGMPDHFDAKMGIPDHMLDALDSAVNMLVTAVDAFTQAGFPAEL 2622
QY 25 ----- 24
Db 2623 LQVHPAQAATQGTGIGMESLAKVFTVRLGSDRPSDILQEBALPNVIAHTMQSLVGG 2682
QY 25 ----- 29
Db 2683 YGSMIHPIGACATAAVSIERSVDKIALGKADLVVAGGIDVQVESLTFGDMARATETKR 2742
QY 30 ----- 31
Db 2743 MTDGIDDRF 2752

```

RESULT 7
 103222
 probable polyketide synthase module 2 - streptomycetes hygroscopicus
 C:Species: Streptomycetes hygroscopicus
 C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Nov-2000
 C:Accession: T03222

R:Ruan, X.; Stassi, D.; Iax, S.; Katz, L.
 Gene 203, 1-9, 1997
 A:Title: A second type-I PKS gene cluster isolated from Streptomycetes hygroscopicus AT
 A:Reference number: Z14848; MUID:98085959; PMID:9426000
 A:Accession: T03222
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1762 <RUA>
 A:Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38062.1; PID:g2624949
 A:Experimental source: ATCC 29253
 C:Superfamily: Streptomycetes hygroscopicus probable polyketide synthase module 4; 3-ox
 C:Keywords: [acyl-carrier-protein] S-malonyltransferase homology
 F:54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:550-822/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
 F:1598-1669/Domain: acyl carrier protein homology <ACP>

Query Match 39.74; Score 62; DB 2; Length 1762;
 Best Local Similarity 2.0%; Pred. No. 7,6e+02;
 Matches 15; Conservative 4; Mismatches 10; Indels 713; Gaps 3;

```

QY 2 RXH-----XHSMEK----- 11
Db 834 RNHNEVRSINELGVESVGSVDPRALPTGRRTSLPTFPQRDITWLAHAGGAAYE 893
QY 12 ----- 11
Db 894 GAGLGTIDHPLLGAVTVADTDELISGRLSTHPLTDLTVNGTVIYPTALIDLALH 953
QY 12 ----- 11
Db 954 AAERTDHTVDELVHTPLALHTPPRSOSTVGAETDGHDPALHSRDTGTWRHTTGT 1013
QY 12 ----- 11
Db 1014 LSNQHPAAELSTWSPTDAHQDITTAQQLADDTGIAHGAFOGLITLXQDNITFAEIE 1073
QY 12 ----- 11
Db 1074 LPEAGLPTGILHPALLDALHPAASANDSEGIOLPFSMAGVTHSTGATHLAVRL 1133
QY 12 ----- 11
Db 1134 DYDDGALIRRLATDTTGQGVITIIISLTTRPLRPDQLAAQDPETHLYLGSPPPIPTT 1193
QY 12 ----- 11
Db 1194 TGHFEKTYVPTAGDDIADTHTITTAQTLRIQHLADTTPLVYEATHDDLAGARVGL 1253
QY 12 ----- 11
Db 1254 IRTAQTEHPGRIILLDTDNHPTSRDALASVVASGRPARIRDGATLTPRLTSLVDTBAS 1313
QY 12 ----- 11
Db 1314 TERRPDLMSGPVLYTGCTGLASLIARHLVTEHGVLDIVSRQSPADAGATELTNOAQ 1373
QY 12 ----- 16
Db 1374 HGARIRITSCDLDRALTLTLDELIGPLGVHTAGALADTTIDHLNDPTLTTLTPKAN 1433
QY 17 PAVY-----TPDIN----- 20
Db 1434 PAWIMHELTDHDLALFVSSVAGVLNAGQANVAAANSFLDALITRRROGLPTGSIA 1493
QY 21 ----- 20
Db 1494 WGMOREGGMTAHLTOADHRTFRAGIHGLTDAEGTTLFTVALDNGLCANAPVKALHPTL 1553
QY 21 ----- 30
Db 1554 NRADTVPAVLRGVAAVAPRAR 1575

```

RESULT 8
 134022 zonadhesin - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C:Accession: T34022
 R:Hardy, D.M.; Garbers, D.L.
 J. Biol. Chem. 270, 26025-26028, 1995
 A:Title: A sperm membrane protein that binds in a species-specific manner to the egg ext
 A:Reference number: 221464; MUID:96064658; PMID:7592795
 A:Accession: T34022
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-2476 <HAR>
 A:Cross-references: EMBL:U40024; NID:g1066465; PID:g1066466; PIDN:AAC48486.1
 A:Experimental source: strain Melshan; testis
 C:Genetics:
 A:Gene: Zan
 C:Function:
 A:Description: may be involved in sperm adhesion to the zona pellucida

Query Match 39.7%; Score 62; DB 2; Length 2476;
 Best Local Similarity 1.6%; Pred. No. 1.5e+03;
 Matches 19; Conservative 2; Mismatches 7; Indels 1133; Gaps 5;

OY 4 HXN-----SMEXR----- 12
 DB 762 HCHRPSSRMCCOTFCGCTHTVCLKNGOYCHPYGSATCYGDPHYLTFTDGRNFNG 821
 OY 13 ----- 12
 DB 822 KCTYTLAQPCGNLTETFEFRVLYKKEERQEGVSCLSKVYTLDESTVTLKGRHTLVGQ 881
 OY 13 ----- 12
 DB 882 RYTLPAIPKSGVFLAPSGRVEVLTAFGLRVNDGQQLFVSPSTFSGKLGCGDYDC 941
 OY 13 ----- 12
 DB 942 DSSNDNKPDPGSPAKDEKELASSWQSEDADQOCENQVSPSCNTALQNTMSGPFCQ 1001
 OY 13 ----- 12
 DB 1002 LVAPHGVFAECPLRLHASSFFKSTPDMCMFOGLQHLCAHMSALTEENCODAGTYKPMR 1061
 OY 13 ----- 12
 DB 1062 GPQFCPLACPRNSRYTLCAHLCPTCHSEPSGHNCKNDRCVGECDPQGVLSGLQCVSR 1121
 OY 13 ----- 12
 DB 1122 ECGCLDSTAGYAVAGERWETFGRCQICBGNKTRCVLMRCAOEFCCGOQDGYGCHAQ 1181
 OY 13 ----- 12
 DB 1182 GSATCYSGDPHLLTDPGALHHTGCTTTLTTPCMLRSLDLSFLVSATNEFRGMLDS 1241
 OY 13 ----- 12
 DB 1242 YRAVAVQVFNLRISLTKGRKVTLDGRVALPLMPAGRSVLTSSGFFLLTYDFGLQYR 1301
 OY 13 ----- PDINP----- 18
 DB 1302 YDGDHLVETVPSSTAGRLCGLCGNTNNNSLDDIQLPDRKPSASSVRLGASWRINELSEP 1361
 OY 19 ----- 18
 DB 1362 GCFAGGKPPRCJGKEVADAKRNCVLMNPGSPSQCNRYVAPSSFSCLYGQCATKG 1421
 OY 19 ----- 18
 DB 1422 DTLTLKSLQAVASLCARAGQALTNNGTFCPLKCPSSGSSVTCANPCPATCLSLNPNPY 1481

OY 19 -----WYX----- 21
 DB 1482 CPSTLPACBCEGCKGHILSGTSCVPLSQCCGCTTQSGSYNPGESHYTNTSCSRCLCTGSA 1541
 OY 22 ----- 21
 DB 1542 HNNISCRQASCKPQCMQDGLIRCVAGMGVCIRIPDTSHTVSPDGSYHAVRGNCTYVL 1601
 OY 22 ----- 21
 DB 1602 VKICHTMDLPFKISGENKREGQPPAFYLRQVYVDIENLVTLKQDQVLLINGTRVSLP 1661
 OY 22 -----XGRIR----- 26
 DB 1662 ATQIGSVARVINDGYTLINIGVQVKNPDGKGFLEVEIPKAYIGRTGCGNPNDEED 1721
 OY 27 ----- 26
 DB 1722 ELMPSPDALLDVNYVDSWRUKEIDPNCQEDDRKTEASQBPQPSANCRPADLEHAGQC 1781
 OY 27 ----- 26
 DB 1782 QAAFOAPMANCATRVVSPYVRSCTHKLCEFGGLNHAFCESLQAFGAQAQGIKPPVW 1841
 OY 27 ----- 26
 DB 1842 RNSFCPLDCSAHSVYTSVPCSPCLPSCQDPBEOCTGAGAPSTICEGCICEPGLVSEQC 1901
 OY 27 -----PVGRF 31
 DB 1902 VARSQCCRDARGTGLPVGRF 1922

RESULT 9
 T14591
 actinomycin synthetase II acmb [imported] - streptomycetes chrysomallus
 C:Species: Streptomycetes chrysomallus
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 03-Nov-2000
 C:Accession: T14591
 R:Schauwecker, F.; Pfennig, F.; Schroeder, W.; Keller, U.
 J. Bacteriol. 180, 2468-2474, 1998
 A:Title: Molecular cloning of the actinomycin synthetase gene cluster from Streptomyces
 A:Reference number: Z18152; MUID:98233744; PMID:9573200
 A:Accession: T14591
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-2611 <SCR>
 A:Cross-references: EMBL:AF047717; NID:g3114611; PID:g3114612; PIDN:AAC38442.1
 C:Genetics:
 A:Gene: acmb
 C:Superfamily: peptide synthetase PPSD; acetate-CoA ligase homology; acyl carrier pro
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:509-966/Domain: acetate-CoA ligase homology <ACLI>
 F:982-1050/Domain: acyl carrier protein homology <ACPI>
 F:1578-2015/Domain: acetate-CoA ligase homology <ACLI2>
 F:2031-2098/Domain: acyl carrier protein homology <ACPI2>
 F:1014-2063/Binding site: phosphopantetheine (Ser) (covalent) *status predicted

Query Match 39.7%; Score 62; DB 2; Length 2611;
 Best Local Similarity 0.8%; Pred. No. 1.6e+03;
 Matches 19; Conservative 5; Mismatches 7; Indels 2346; Gaps 6;

OY 1 SRXN----- 4
 DB 116 AREHLEFDMALLRLGDDRCXWQGYHMYMDAFGYLITRTAEVYTTALASRPVSPSPFG 175
 OY 5 ----- 4
 DB 176 SLDFIADQKTRDSEGTGDHAYTERFADRPETGIVANPSTTPDHLYLRTALPGE 235
 OY 5 ----- 4

Db 236 LDGKFAARGARAPWSHVIYAAVYLIHRMTGATDVVIGLPTVATRLDQLOKRTPTGASNV 295
 QY 5 ----- 4
 Db 296 LPLRLTVRPELTLKOLLTOVSARIVELGRHORTRAEDLQDLDLPGGLGTWYAPVNNIMS 335
 QY 5 -----XISM----- 8
 Db 356 FDATPTFAGLPTTAHSLSSGLVBDLTFPAMDNRDAGLJTVDLNAHPELCPERELTAHGR 415
 QY 9 ----- 8
 Db 416 LVATLRAIASDIDRPYGRIDLTLAEERKALLAGPEDAPATGATTEATLPELFRANTAA 475
 QY 9 ----- 8
 Db 476 APDAVAVVCDSETSLTYREIDERANKLAHLAAGVGPERYVALALPNSVDLVAVAVLAVLK 535
 QY 9 -----EXRTPD----- 14
 Db 536 AGAAVPLDPEYRANRLAHVYTAQPTLLITTTETEAKLPDRHAPALRLDDPETILALA 595
 QY 15 ----- 14
 Db 596 GOPANSPAVGLRDPHAYVITSGSGVPGVNTHRNVRLFDATPWFDFGPDVWTL 655
 QY 15 ----- 14
 Db 656 FHSTAFPSVWELNGALLHGRVYVPIVDSRSPHATLIDLADOGVYVLAQPSAFHOLA 715
 QY 15 -----INPA--WT----- 20
 Db 716 QAAADGPRPRRLARVYVFGSEALQPARLAEMYRRHEDTPOLVNMTGITETVHTHQ 775
 QY 21 ----- 20
 Db 776 PLTRDRAAGAASYIGAGISDLRTHVLDGLQVLPAGVSELVYVAGPLANGYKGRALT 835
 QY 21 ----- 20
 Db 836 AERFADPYGARGARMTGDLVVRNPDGELFEGRADHQQVKGFIELGEVEAALLAH 895
 QY 21 ----- 20
 Db 896 PDVBOATVYREDRPGDRLVAVYVGREALRPEQVRETFREKRLPEHNPVAVVOLERLPL 955
 QY 21 ----- 20
 Db 956 TPNGKIDRALPEPDFFALAGTGREATPQEQIVCDLFTQVLGLPRVGVDDOFEELGGHSL 1015
 QY 21 ----- 20
 Db 1016 LATRLIARLRTVLGELELRSLFSGPTPAVAARLDTAGORLALTVRQRPVPMPLSSAQ 1075
 QY 21 ----- 20
 Db 1076 RRLMFLSTLSEPSATYNIPLVRLSGRLDYPALCALGDVGRHESLKTVEPEVDGTPYQ 1135
 QY 21 ----- 20
 Db 1136 RVLTPBAAPRLVTPPSEADLPNALKAGARIANDLAGEPPLRTLELSEPREHYLLVVA 1195
 QY 21 ----- 20
 Db 1196 HHIAADGMSHPLSRLELEAFAARAGRAPQAPLPQVADYTLQNONELGDQNDPDLF 1255
 Db 1256 ATQVAVTETLAGLPDQITLPTDRPRPAVMTYRGDYLVVDIDPELHRLTELANGSGASL 1315
 QY 21 ----- 20
 Db 1316 FNVTOAGIAALLKRLGAGDDIPLGSPLAGRTDQALDLDLIGFVNTLVKRTDTSDDPTFTQ 1375

QY 21 ----- 20
 Db 1376 LCTRREHSLAAVADVPPEVLEVINPTRLAHNPLOQIMIALQNAPENEFSLPGLRA 1435
 QY 21 ----- 20
 Db 1436 GIELGTGAKFDLPSLTERGPDGEPQGLTEVEYSSDYEATVOALFDRLVRLHA 1495
 QY 21 ----- 20
 Db 1496 VTAQBPQLSRIDVLTPEERNRTVEVNRTELPLDASIAELFEQOVTLPDAPALYSDG 1555
 QY 21 -----XKGRP----- 27
 Db 1556 ATLSYSELNTPANHLAHQLTTRGIRPGAVAVILORSPDTVTVLAKTGATVYIPLDSR 1615
 QY 28 ----- 27
 Db 1616 YPADRIYVLDSTRUKLITDHTIDLTITTOFPNADTPHDEDPGNHPTHPDAAVI 1675
 QY 28 ----- 27
 Db 1676 MTTSGSTGRPKGVIAHNRITLALDPRDPTAHRRVILHSTAPDASTYEIWPPLNGN 1735
 QY 28 ----- 27
 Db 1736 TVYLAPTGDLVYVYHRVITDQITAVFLTALFNLTEHPACLAGREVWVGGEAVSA 1795
 QY 28 ----- 27
 Db 1796 FSVRVQEACPSVWVDVYPTETTFATHNPVPPYTGPAVVALGRMATMAVYLDQA 1855
 QY 28 ----- 27
 Db 1856 LDPVANGVVELYLGAGLAGYLDRLPALTAERFVANRBERMRTGDLAKMSADAQL 1915
 QY 28 ----- 27
 Db 1916 EFVGRADQOVKVRFRIBEGELENVLTGHPAVAQAALLVREDOGPRLVAVVADGTA 1975
 QY 28 ----- 27
 Db 1976 PDGLREALRRRLPEYKWPSTAFVYLDRLPLTANGKVDRAALPAPVWSRAGSVREPTPBG 2035
 QY 28 ----- 27
 Db 2036 LLDGLEAQLGVRNGLYEDFFALGSDSTVSTRVLSARSAGTGSVRYVFEQRTAAGLA 2095
 QY 28 ----- 27
 Db 2096 VVAEELTASDGEADESAGVPVTPPIRMWFERGGSIDRFHOAMLLQVPALGEXRLVYA 2155
 QY 28 ----- 27
 Db 2156 VQALLDHDALRLTRVPSSEDDGDPWELHTPAAGTISAADLYRVRODLDPDEATVTHL 2215
 QY 28 ----- 27
 Db 2216 DAAQARLHAERGVLLQAVVFDAGPRRGRLLLVNMLVYDGVSWRLLPLDVAAMEAVTAQ 2275
 QY 28 ----- 27
 Db 2276 GDPRLPEVPTSPGRWSRLAAEARPRABAALAEALMTVEVLPVDPPLATRPDRTRTVG 2335
 QY 28 ----- 27
 Db 2336 RANSLTVSLPEEVSPLLTTPAAPHAGVNDVLLTALALAVARRGRGGRPHTALLDV 2395
 QY 28 ----- 27
 Db 2396 EHGREGIYDGTDLSTRVGNFTSLYPVRLDPGALGWEVYTGAPALGPAKRVKBOJREL 2455

QY 28 -----VGRF 31
 Db 2456 PDRGIGRLRLHAPWGPVLAQIATPOLGPNILGRF 2492
 RESULT 10
 A35548
 319k protein ndv8 - Rhizobium meliloti
 C:Species: Rhizobium meliloti
 C:Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 08-Oct-1999
 C:Accession: A35548
 R:IdP: L.; Dylam, T.; Ditta, G.S.; Helinski, D.R.; Stanfield, S.W.
 J. Biol. Chem. 265, 2843-2851, 1990
 A:Title: The ndv8 locus of Rhizobium meliloti encodes a 319-kDa protein involved in the
 A:Reference number: A35548; MID:90153914; PMID:2154461
 A:Accession: A35548
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2870 <ID>
 A:Cross-references: GB:J05219; NID:q152270; PIDN:AAA6305.1; PID:q152271
 C:Keywords: transmembrane protein

Query Match 39.7%; Score 62; DB 2; Length 2870;
 Best Local Similarity 0.8%; Pred. No. 1.9e+03;
 Matches 14; Conservative 4; Mismatches 10; Indels 1694; Gaps 3;

QY 4 HXSMEXR----- 11
 Db 180 HTSTVRESITAMVGFQHEETIKIGELNALPSILRFVLIENIRLAIWERSRGRRK 239
 QY 12 ----- 11
 Db 240 ANEVAQDLRLNDEGCRITLVESEALADNTIAQLTYMRDGSOSSGAVIAMIERLE 299
 QY 12 ----- 11
 Db 300 RRGDVEALVAQNRLSGNATWSNIRSLREIDTDNAWFEVSXKIDATREGSDYA 359
 QY 12 ----- 11
 Db 360 ALPGSNRYRTYETIKARRSGHEHEVETIAEWEBAKAAVAEAPLOEPNVSFLVQ 419
 QY 12 ----- 11
 Db 420 KORLALEKRIGSPSLFOHLRSVRKIDWFAIAGPNILLTILAMIVYAFVSPMDIPSGA 479
 QY 12 ----- 11
 Db 480 KLIMLLFALPASEGANGLFTVFTLFAKPSRLVGYEFLDGIPEARTLVVPCILAKRD 539
 QY 12 ----- 11
 Db 540 HDELYRNLFEVHYLANPGRGIIYFALLSDMADSKSEAPADTVLEYAKRBIASLSARVAY 599
 QY 12 ----- 11
 Db 600 DGKTRFELHRRRLYNEAGVWGMGKRGKLEHMLNLLRGDNDTSFLQANNVPSGVQY 659
 QY 12 -----TPDIPA----- 18
 Db 660 VMTLSDTRLMRDVATKLGKLYHPINRPVYNERTQEVYGYSLQPRVTPSLTGSSEAS 719
 QY 19 ----- 18
 Db 720 AFGRITINGIDIPYFTVSDVYQDLAGSGSFLGKLYHVDFAEALKRIENAVLSHD 779
 QY 19 ----- 18
 Db 780 ILREGSARCALVTDLVEDEPIRYEVSRSQRHARAGDQMLPYLPNFKGLSMLGRKK 839
 QY 19 ----- 18
 Db 840 MYDNLRSLLIPVAMLAASVGVTKMEPTPALIQWLVLFSLVAPTLSDISGIMPRRNDI 899

QY 19 ----- 18
 Db 900 VARAHLFTVLSDIRANQVALRIVEIAHNAAMADATVRSIXRTFVSKRLMEWTAQ 959
 QY 19 ----- 18
 Db 960 VDSAGSGTGDYFRAMWTAPALVALSLAALSDTGLPFTGLPFALLMAASPAVAMFVSQ 1019
 QY 19 ----- 18
 Db 1020 SAFTEDQVYSEEAIEEKRIARTRTRTEAFVTAQNTLPDNGQETPOVLAERTSPT 1079
 QY 19 -----WYXRGIRP----- 27
 Db 1080 NLGVYLLSVMSARSPGWTGFEETITRLQTIATIDRMPEYRCHLFNWTRTGSLPEMEPRY 1139
 QY 28 ----- 27
 Db 1140 VSSVDSGNLAGHLIAYSSKCREMAEAPSAHYQNLGIGDVAAILKEALINELPDDRKTVR 1199
 QY 28 ----- 27
 Db 1200 PLRLVEERINAGFQNALAAVKEEREFASIRVINTLAVIARDMKLTVINLDEHYVQSGEV 1259
 QY 28 ----- 27
 Db 1260 AYWAGSLVACBAHINDGVFDGATEALRQRLVLKERARDIAFSMDSFLRPERRLLS 1319
 QY 28 ----- 27
 Db 1320 IGRVYANLEDACTDYLASEARLTSLFALAKGLPTEHWYKLGRIPIVIGARGALVWS 1379
 QY 28 ----- 27
 Db 1380 GSMFEYIAMPVYMRQGGILNQTNNIVYQEQINHGRLGTFWGISSEAFNDEHETLYQ 1439
 QY 28 ----- 27
 Db 1440 YTFNGVPTLGLKRGIGNAVIAPYASILLACMDPKSALANLRLREVGALGAYGHDAVD 1499
 QY 28 ----- 27
 Db 1500 FTPTRPBGQCAVVRANTYAHHRGKSVAAVANVFNGQLREWFHADPVIEAEELLQEKA 1559
 QY 28 ----- 27
 Db 1560 PRDIPVAAKREPEALGKGADLLRPEVRVYEDPIINDRETYILSNGHYSMLTATGAGY 1619
 QY 28 ----- 27
 Db 1620 ARMGOSVTRMTDPVEDRTGTFIFLRDVTGDMWSATAEPRRAGEKVTYRFQDKAEF 1679
 QY 28 ----- 27
 Db 1680 VKTVGDLTSVEVCIVATEHDAEGRVILLNTGTEDRTLEVTSYAEVYLAADADSSHPTF 1739
 QY 28 ----- 27
 Db 1740 SKMFLTEISRHDVIVYVSRKRSRGPDIYAVLVTDNAGSERHTQETDRRRFLGGR 1799
 QY 28 ----- 27
 Db 1800 TIAEAAPDPGATLSGTGFTLIDPIVSLRRVVRPAGKRVSVFTWTIAAPDRREGVDRAID 1859
 QY 28 -----VGRF 31
 Db 1860 RYRHPETFNHLEIHWTRSYQVMRHWGITSKEAASFOMLGRY 1901
 RESULT 11
 T12117
 polyprotein - fava bean dsRNA replicon

C:Species: *Vicia faba* (fava bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: F12117
 R:Referrer: P.
 J. Gen. Virol. 79, 2349-2358, 1998
 A:Title: Nucleotide sequence, genetic organization and expression strategy of the double
 A:Reference number: 217424; MUID:96451313; PMID:9780039
 A:Accession: F12117
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-5825 <PEP>
 A:Cross-references: EMBL:AJ000929; NID:g3184155; PIDN:CA04392.1; PID:g3184156
 A:Experimental source: VILION; cultivar 447
 C:Comment: This gene product may be cleaved into several proteins including helicase and
 C:Genetics:
 A:Genome: dsRNA replicon
 C:Superfamily: fava bean dsRNA replicon polyprotein

Query Match 39.7%, Score 62, DB 2, Length 5825;
 Best Local Similarity 0.4%; Pred. No. 7.4e+03;
 Matches 17; Conservative 1; Mismatches 10; Indels 4154; Gaps 4;

QY	4	HXHSXEXT	12
DB	1336	HAHSEAGTSCCLIPAGEACPCGYNROIESEILENSEESADEAEONRVKNSKP	1395
QY	13	-----	12
DB	1396	RHPDEKGNRSNNRSDNRDANHSNVNHHGNKPKQOGATQPPKRSALNPEND	1455
QY	13	-----	12
DB	1456	NDTPSTPTVILTADPTNPGEQAPGDNIPIENNDIPGSTTQSSPPDDINYSGE	1515
QY	13	-----	12
DB	1516	NNRNTONRIRILIACTNEPFGDLAYGNDLAVKNSKMLAPPNEORLSCFOFNMGDSI	1575
QY	13	-----	12
DB	1576	YATENLEVIKTYNFGEGGYGNLKLIVPLDLTEMQEIVSETQFSDWEIMRYAQ	1635
QY	13	-----	12
DB	1636	AKQNLIVTERCALVNSVCSNEPGVICHNRNGVMEHEMALAIQGFADYHPFTN	1695
QY	13	-----	17
DB	1696	ALTREDLIDPAKSSGLNKKVHSFVLDPRILAKIOTEMHESLSKVEDVNAFGVAKITKG	1755
QY	18	-----	17
DB	1756	STHYTNDNNNLGYSGTGTAAJTEDNQELVELLLAASAPANVITDWFYDRPESP	1815
QY	18	-----	17
DB	1816	ADATGINDYGRQVDAEVAQACADLISELTONRYDCKEYINNHLKVOYKTKLK	1875
QY	18	-----	17
DB	1876	PGDLISGLVGIQSTSYDSHVIGVIRTNQIMCTCYEKAKAKIKINDLXKRTGSGL	1935
QY	18	-----	17
DB	1936	RALFGIFRMDPRDIKTLLEKATAVDAIAGMGRSTEIVKLVNODCTVAQTSAAVSNTLE	1995
QY	18	-----	17
DB	1996	KLEGGKNNKVMSEKCMTOOVNPTPLVIDEASMTWETLSLTGPQVENVLYGNTL	2055
QY	18	-----	17
DB	2056	QICVLDNRTGGSRAVTKSLIQAGIIRNRYTTHRIGNPLANELSVYKEILTNAKNETNF	2115

QY	18	-----	17
DB	2116	CTKSMDAVRMAELTSLAGSLEPVLCTYNNAVRAVMTLKVGCRVDIINKFOGLEADNV	2175
QY	18	-----	17
DB	2176	IVLQWCPGTGTPGRITLDRNQCSAATRAKMLWISVNEYSNNVPLNRMGATIGSKH	2235
QY	18	-----	17
DB	2236	TOPDTENNLADKTLQVYVSHLSKVMYTSKDKTIERLECSRLOSQEBGNLYMLKNTLHLD	2295
QY	18	-----	17
DB	2296	VDYSGNDQSLRQSGKLNKFKSNLSFSKPAADVSSSTLGLAEIVQANAND	2355
QY	18	-----	17
DB	2356	RINSQINNTYVOKNLAISNGSIAAATVPVLTAMLAQRPGLTKACSPDEPTETLVANI	2415
QY	18	-----	17
DB	2416	IMLGAVNLKIQDPRNDKVLSSGGLRSFYADSSETKOLYNOKFKCCGHVYLKYNPDTE	2475
QY	18	-----	17
DB	2476	ISVKKKMFERVLIVNPGVCLSGSLHSVTIENELKMQSYGCSLCGAVTEHNNK	2535
QY	18	-----	17
DB	2536	PIVGDPSYNTGNSRTVFNKSKQDCKPILAVLRMLNVDGTVKLPYDPVGRGDFDM	2595
QY	18	-----	17
DB	2596	PEGDNVAAALVIRVNLKGPMAAAKLNMLTNTNODLCFNTDNNSELPDSIKARLESNG	2655
QY	18	-----	17
DB	2656	FTNENTNLVEFVGRTLEFNGFTNNTTELVNNSMDLCRPICTLDKRYLAALMDQDN	2715
QY	18	-----	17
DB	2716	LIVTQISGAPIHPPAGVTVSNKAEKTMAYROMLNLEFKLTNTQLSDKPKGAPNPG	2775
QY	18	-----	17
DB	2776	PLKFMQHEADNDKVAVAAKRYKGTQAAAMKLSRRKYLITAMVNGNSEYASNLMOQH	2835
QY	18	-----	17
DB	2836	NVITTSINYPENMLGLVDLYVAQAINNTGYTSALYITNNACTAVLGHMDWFSPPDGG	2895
QY	18	-----	17
DB	2896	WSTSRFLSTHTTTNLADIKSLDSEBLAKDROELSEETRAVLKKEKRSERJHNOIL	2955
QY	18	-----	17
DB	2956	RTEDSWYSESQVAKRGTVIVSVNSGISTESTIKIMETTGAMEFLMCYPTLPSDNTTA	3015
QY	18	-----	17
DB	3016	HRVGTGSNNVKITTPGARNLTLPNEELIYCMKSGRAVQOSQTKKDFMISLNTTVAM	3075
QY	18	-----	17
DB	3076	IASVQIKATTTPRNEHAKPLIPSPSPSNOSDIKVVAVQTPRTCPNGDSNSLII	3135
QY	18	-----	17
DB	3136	TKRRLVSKALLAKMNERALRHSNLEKDAQYGRSLTLMQVYTERVNVYINKNGDLEEA	3195

```

OY 18 ----- 17
Db 3196 VCIALLTKHRIINOLEHTMGLMETFRYVDTSITDIYVSCKANLEKVMSTANWQEK 3255
OY 18 ----- 17
Db 3256 LGIPASFTGLDELANTORVVFIDERTNOSVMSDNSIKHMSDEBSIKKWASTADNL 3315
OY 18 ----- 17
Db 3316 DMANIGMIOEYRKVLCREDIPKTKACPGFKQYVDCAKAQAOKKARKTDPGARVLYLE 3375
OY 18 ----- 17
Db 3376 EPIDLIGNVDITRAOTGTWTKNENIRSEITFRANQSIIRAOQTSIRSTLDMNEMVK 3435
OY 18 ----- 17
Db 3436 AYSALBAGDLKVMYBELMSRIAIKSTSTPMHLKEDLEMDVFNLTREHLEITKLIG 3495
OY 18 ----- 17
Db 3496 DAVVSRMOSKAEKQNEYDFIKSDIGLSTLNDHDLNIVLSLAVDRPSVKKDLTNL 3555
OY 18 ----- 17
Db 3556 GSOXSVNEAVMLRKIKNOOCESASNGQLVDHDMANSVSTEONOGKLNILSLHEEA 3615
OY 18 ----- 17
Db 3616 GICDINSDASPLKIRASSTLEAITTCVOTPGMYELTSPKEIITYAMNDECVARCIE 3675
OY 18 ----- 17
Db 3676 KYVTONIEPNRITNGLRATMOOSKMLTESQALIVCOLLGNCLLIONGETGVYNAPN 3735
OY 18 ----- 17
Db 3736 KPPVQAKRLSODASIDHCYLINLIGADGVKRLSPENIAKENMOLEHVCVDNKPPIYAVG 3795
OY 18 ----- 17
Db 3796 ENFYACVSHVHIDNEDLARLSCSQPYEKIPYLDGREMLPVNNYMARMRKTSATLAK 3855
OY 18 ----- 17
Db 3856 KVPITTTAGSLALTRNYKATIPMYTKPGDVLVYLRDCEKRYMCTHTTIVDPGGEVQ 3915
OY 18 ----- 17
Db 3916 ILETITDIDGAVCLDGLHCLTPADRTPRDRPVNTSNEVMSLINTORTYLEATYGLRS 3975
OY 18 ----- 17
Db 3976 HVGQGEKLNHYHONREHMYDENDYIDAMLDENIYIELDPTVOYPTGVNIYNLICIR 4035
OY 18 ----- 17
Db 4036 MFLAMLXGIPYVAVCDNPSAQAIAWFLKYYKMGHVENNNVTTEDLNTRAMEIINA 4095
OY 18 ----- 17
Db 4096 LESTYQMTTNNTRVHLRKAETILSKKEVSEELPNTTIDISVTAVLMGLENVIAKSN 4155
OY 18 ----- 17
Db 4156 QGTCEVFLIPETEPOLADOMLESKGLIYDIKOCANVHYPTGTVALRYVESLSNEIAC 4215
OY 18 ----- 17
Db 4216 STAVKGVDPDDEKRAQASNIQSVPASDKADGSSOLSCEBMSQOTGTQVACVHIEB 4275
OY 18 ----- 17

```

```

Db 4276 CDEANASNDVYLKETTOEONNHDIISKILATQAQAKSNPBDKSWKSGATSNADQATOS 4335
OY 18 ----- 17
Db 4336 LCDSPDGMSOVADOLAKLELVNFSDEYTLGHIORNYVNSOGTYLORVDISERWALT 4395
OY 18 ----- 17
Db 4396 YNGIGHGSDVPPIINPGMWMLAEPPIKGLGYKKEHEIVYINAGNCAYRLNRIKPREP 4455
OY 18 ----- 17
Db 4456 KSVFTNNVREVALDILTYTSANNOQYPTNLOBLDKLHGVSWMVYGEQYNDPLQT 4515
OY 18 ----- 17
Db 4516 CVELTRMERVYVTDAPVYNQDADIQREKIITIKVTSOPEEYRAIMWRMILFANI 4575
OY 18 ----- 17
Db 4576 DPCELIYILNGREMAPNDMEIVDRWESFEDPDIVSALHCAFHVCCGYAKYNARFTD 4635
OY 18 ----- 24
Db 4636 YNTFNDIYRAGNSYGDIEIYNSRYKMMMLYIILPGGAMYIGSKMLNLTDTTICLE 4695
OY 25 ----- 24
Db 4696 LKGEETRIDHYKGLFCFSANFHVADPTGIYDSTVMSVDGIPMOYAMOJINSHVTE 4755
OY 25 ----- 24
Db 4756 NPYRHSIGESTGNPRNISDGEKPMQANNNMLDPLGLKSHKMIQKFLDSSVAKIK 4815
OY 25 ----- 24
Db 4816 ESQSLEPIOLGPKPIYVFANQCEAIGVSMNNTGILYVEVINAQOKLAARNEPLOHTLI 4875
OY 25 ----- 24
Db 4876 LKSHLTLMSTGGRHSWKREYTHLAKADLONCOYVEGASISLQWKRETOHACVYTE 4935
OY 25 ----- 24
Db 4936 ERDVFYRIPQALDHSVGLTKLEVYTGNEGLSSSGYLARVLRKINDRLAMNHLTOH 4995
OY 25 ----- 24
Db 4996 LOKLEFDRILISHESSGATFKPRIGOTIYLAADVLAAGIECVNHIITTSREKKAALK 5055
OY 25 ----- 24
Db 5056 NPLTKYGINPLNSVIVSGLGCVOPTIPRCYAMLCQYKLIKPEPTRGGGTPSNVSAN 5115
OY 25 ----- 24
Db 5116 VGPSNPTTAGPDTPPLITSNKNVESGASNSQSPGSAKAYQDKLRBKLETPODLRY 5175
OY 25 ----- 29
Db 5176 LETMLEDAVALMSGLTDLRLNDGWNHCAIDATMKSDPEIIEPVGDILLPLSNTSAL 5235
OY 30 ----- 29
Db 5236 RHNFTNARYIDLMDDTDLADMLTLAPRNPMLTSRVSPPGQKINLKTLLNRPCCORPV 5295
OY 30 ----- 29
Db 5296 PTQVAMGENAVTGRIGSVLPLRREPMNVTHELHKFRTAYYRDMGMEVLRKDFRANITIS 5355
OY 30 ----- 29

```

Db 5356 DADVATLSRRSDWKALATSTIKMETGLPSNPMAVAVHVKTESLTKAMPIMWRQTOG 5415
 QY 30 ----- 29
 Db 5416 RIIVQPELCAALSPAFIAIKRLEKLEIVYDGLAPPDLSARAPTIQDYVEED 5475
 QY 30 ----- 31
 Db 5476 LVIDRQRTDQELIDLEFQVMDLGLDINAMIMRLVHKRPF 5517
 RESULT 12
 T30226
 polyketide synthase - Streptomyces hygroscopicus
 C:Species: Streptomyces hygroscopicus
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T30226
 R:Apuricio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
 Gene 169, 9-16, 1996
 A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
 A:Reference number: 220782; MUID:96186896; PMID:8635756
 A:Accession: T30226
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-8563 <APA>
 C:Cross-references: EMBL:X86780; NID:9987088; PID:9987100; PIDN:CAA60460.1
 C:Genetics:
 A:Gene: rapA
 C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:54-509/Domain: acetate-CoA ligase homology <ACL>
 F:1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:3199-3270/Domain: acyl carrier protein homology <ACP1>
 F:3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:4787-4858/Domain: acyl carrier protein homology <ACP2>
 F:4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F:5366-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F:5760-6831/Domain: acyl carrier protein homology <ACP3>
 F:6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
 F:7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
 F:8412-8483/Domain: acyl carrier protein homology <ACP4>
 F:4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 39.7%; Score 62; DB 2; Length 8563;
 Best Local Similarity 0.3%; Pred. No. 1.5e+04;
 Matches 19; Conservative 3; Mismatches 6; Indels 7557; Gaps 6;

QY 4 HXHSKEX----- 10
 | | : |
 Db 420 HPEHEVALDVPGVKDAAGKQHPVLGETPVLYVETGVDVNLAVCERLSYFK 479
 QY 11 ----- 10
 Db 480 VEEIYRDALIRASGVKNSLTKEPAELLAGSGGETLHREWIPLDPPKQASPDGR 539
 QY 11 ----- 10
 Db 540 HVVYEVDSLPDDADLAGAVNDLAQSWLADKRRADSTLVFTYBRAVHTGPSDLVPEHNA 599
 QY 11 ----- 10
 Db 600 AMDAIRREQTENPGVFFVVDVDDADPDPTLLRALAGLEQIALRDGNPLVPRILAH 659
 QY 11 ----- 10
 Db 660 ANTADSDSLTIPEDAMLLHSRSGTLNDLALVPADTERPLRPGEVNIAVRAAGLNRD 719
 QY 11 ----- 10
 Db 720 VIALGTYPGEGLMGEAAGVVLEVPYVHDLTPGDRVFGVLSAFGVAITDRLLGAI 779
 QY 11 ----- 10

Db 780 PDFWSTTAASIPITVFATAVYGLVDLSGLSAGETVLIHATAGVGMAATQASRHLGPR 839
 QY 11 ----- 10
 Db 840 HATASAKOHILNEAGLEDTRIADSTFLAREAFINTTGGQDVVNLNSGDFVDSLND 899
 QY 11 ----- 10
 Db 900 LTRGGRELEMGKTDIRDADRTADRPQTQAFLDLGPPDRIRRIIAELLEPAQCVL 959
 QY 11 ----- 10
 Db 960 RPLPVLTWDIRKARDAFWMSRARHTGKITFTIPRLDPDGYVLINDAGALFTVABHL 1019
 QY 11 ----- 10
 Db 1020 VADQVRIHLLLSRSPDEALINELISGARVDTAVCDVSDRAGLVRTLAGIAPERPLTA 1079
 QY 11 ----- 10
 Db 1080 VHTGGPVAHESHQHLRLTKGLDLAAFYVFGQDAPASVDALASRRRAGLPITAIANGI 1139
 QY 11 ----- 10
 Db 1140 PEAEAVVVRGPLLGRAMASDSAHITVRLNTVGLRALAADILPPLQLVGAHHTDEEQ 1199
 QY 11 ----- 10
 Db 1200 QAWSRQFLAEARREGALNDLVNSVTDILGLSADRVAPDKT SRMGIDSLTSVELRNS 1259
 QY 11 ----- 10
 Db 1260 LKATGRLPLATLVFDYPPPAVIVVRLGELTGESPAPEAVSAVQGGPELAVGMACHL 1319
 QY 11 ----- 10
 Db 1320 PGVSPEDLRLVRESGTDALSGFPTRGDVDGLDPDPDASGKSYCVQGGFLDPAAGF 1379
 QY 11 ----- 10
 Db 1380 DASFGISPRALAMPQCNLVLEVSEAFERAGTERPGSVRGSDTGVFGGPGGIGAGA 1439
 QY 11 ----- 10
 Db 1440 DLGFCATAGAAVLSGRVYFPLGEPALTVDPTACSSILVALHQGYALRQEGCSIALY 1499
 QY 11 ----- 10
 Db 1500 GGYTWFPTPOSEVEFSRQGLSADGRCKAYADAADGETENAGVGLIVERLSDAQAKHQ 1559
 QY 11 ----- 10
 Db 1560 VLAIVRGSANVQDASNGLSAPNPGSQKRVIRALSNAGLAHIEVYVLAHGTTLGDP 1619
 QY 11 ----- 10
 Db 1620 IEQAQVATATYGQDREGVILGSLKSNIGHAAGVSVIKVMALRHGFVPTLHVDP 1679
 QY 11 ----- 10
 Db 1680 SRHYDSAGAVEIYAENRSPATGRPRAGVSAFGVSTNAHVILEGAPASVDYDAGST 1739
 QY 11 ----- 10
 Db 1740 PYVDSVPLVYSAKSLPALAEVBGRKRAYLAASPGADVRAVGSITLVYRSLEFHRAYLL 1799
 QY 11 ----- 10
 Db 1800 GDSVTGTGTAVDPRVYVYFPGQGWQMLGMSALRTSSMVFARBAECAAALSEFVWD 1859
 QY 11 ----- 10

```

Db 1860 LEAVDDPAAVAVDVQVQASNAVAVSLAAVQAGVPPDAVVGHSOGELAACVAGAVS 1919
QY 11 ----- 10
Db 1920 LRDAARVTLRSOVLARGLAGAGMAVSLPQDVDELVDGAWAARNGPASTVVAAGAPEA 1979
QY 11 ----- 10
Db 1980 VDRVLAHVGANGVRRRLADVASHPHVELIRDELLGVLAVGDSRAPVVPMLSTVDGTV 2039
QY 11 ----- 10
Db 2040 VEGPLDAEYWRNLREPVGFEPAGOLQOAGDIYFVEVSASPYLLQAMDDVTVATLRR 2099
QY 11 ----- 10
Db 2100 DDGDAIRMLTALAOAYEGVVDMPAVLGTTAARVLDLPYAFQHORVTLKGVRAAADG 2159
QY 11 ----- 10
Db 2160 HPLGTVAALFSGDVVLTGRVSLATIHMLADRAVRSVLLPGTAFVELVVRAGEVCD 2219
QY 11 ----- 10
Db 2220 VVDELVIETPLLPQTGGVQWSVSAEADSGHRAVAVFSQADNTDTWTHVATVSTSD 2279
QY 11 ----- 10
Db 2280 STVSOPFAWPPAGAPLDLSDFDLTGAGYEGPAFQGLTAWRDNVFAEVALA 2339
QY 11 ----- 10
Db 2340 EEOQOEARFVRHALLDALHASNSTLDTAEOGVLPFSNOVYVHATGPVLRVAIT 2399
QY 11 ----- 10
Db 2400 RTADGSLVAVDGSRPVAVSLVTRVYVADALGSADELMLTWMELIAPQGTGLTVG 2459
QY 11 ----- 10
Db 2460 RFEDLVSGGVPEVVEVFTALPGSSETPLDPLDPLAQTTLTAQVLAQAVQAWLAGEGFT 2519
QY 11 ----- 10
Db 2520 NSTLVYRTGGLAAAGVSGLRVSQSHRPGFVLVECDDTLTPDQLAATVGLDEPLRV 2579
QY 11 ----- 10
Db 2580 CDGRFEVPRLARANTPESPLTIPDSRAHMLEOPRSGTLHDALVPTDAERPLQSGEVR 2639
QY 11 ----- 10
Db 2640 VDVRAAGLNRDYYVALGNVDKRLAGGAGVYLEVGPVODLAPGDVFGVGGFGR 2699
QY 11 ----- 10
Db 2700 SALADRMGLVLPDGSFTTAASVPVFAATAYGLVDIAGLSAGESVLHAAAGVGMMA 2759
QY 11 ----- 10
Db 2760 TQIARHLGARIVATASTGRHILREAGLEBTHIADSKTILSFQETFLNTHGQGVVANS 2819
QY 11 ----- 10
Db 2820 LSGDPTDASIDLPLRGGRFLEMGKTDLRDPHQVYADRPGTYOATFDLADAGPDLREIIT 2879
QY 11 ----- 10
Db 2880 ELTLFTQGVILPLPYOANDIRQARDAFSMSRARIKIVLTIIPRPDPDGIILTGGSS 2939
QY 11 ----- 10
Db 2940 GVLAGILARHLAAHGARRHLLLSRTTPQALIKELAEIAGAVDTATCDVSDRAGIARVL 2999

```

```

QY 11 ----- 10
Db 3000 AGVSPHEPLTAVILHTAGALDDGVESLITQOULTVLRPRADGAWHLHELTONTDLAAFYA 3059
QY 11 ----- 10
Db 3060 YSSAGVLGSAGOGNYAANAFVDALAEORRGGLPALAVANGIMEDTSGLRKMTDTR 3119
QY 11 ----- 10
Db 3120 DRLRSGRLAISAGRGMLDASRGEPLYLAASKEPVADYFPALLRLHRPARRAA 3179
QY 11 ----- 10
Db 3180 STGSSVQNTARLAPYERREKLLKVCDAAYVLGHADSTIPATGAFDGLVDSTAVE 3239
QY 11 ----- 10
Db 3240 LRNGLAKATGILRLPATLVFEDYPTPALAARLEELFTGENPAPVRSVSVAQDEPLAVG 3299
QY 11 ----- 10
Db 3300 MACRLPGVSSPDLRLLESGLDAVSFPTDRGMDVENLFGPAYGNSYRLQGGFLDAAA 3359
QY 11 ----- 10
Db 3360 GFDAEFFGISPREALANDPQQRVLEVSWEAFERAGIKRGSRGIDTGVFMGATYFGTGI 3419
QY 11 ----- 10
Db 3420 GADLGFGGTAGAVSVLSGRVSEFEGEPATVDTACSSSLVALHQAGYALRQECSLA 3479
QY 11 ----- 10
Db 3480 LVGGTVYVATPOTFVFAHQGLIAGDGSKAPADSADAGFSEGVVLVERLSDARRNG 3539
QY 11 ----- 10
Db 3540 HQVLAVRGSAYNODGASNGLAPNGPSQQRVIRAAISNAGLSTAEVDVVEAHGTGTLG 3599
QY 11 ----- 10
Db 3600 DPLEAQALMATYGGDREOPLLLGSVKSMLGHTQAAAGSVIKVYNALORGFVPTLHVD 3659
QY 11 ----- 10
Db 3660 EPSRHVDSACAQVLTENQPPGTDPRRAGVSSFGISGTNAHVLLESALPTOPAGNTV 3719
QY 11 ----- 10
Db 3720 VESAPFVPLVTSARTQSALAEYEGRLRAYLAASPGADTRAVASTLANTRSVEYRAVLI 3779
QY 11 ----- 10
Db 3780 GDDVTGTATNDPRVYFVFPQGSQRAGHGEELAAFPVFAIRHQVMDLIDVDDLVDNE 3839
QY 11 ----- 10
Db 3840 TGYAQPALFALQVLFGLLESNGVRPDAVVGHSVCELAAGTVSGLMSLEDACTLVSARAR 3899
QY 11 ----- 10
Db 3900 LMQALPAGVYAAVPSDEAKAVLGEVEILAVNGPSVYLSGDEAAVLQAAEGLGKWT 3959
QY 11 ----- 15
Db 3960 RLPTSHASHANMEPLLEFRAVABGLTYRTPQVMAAGQVMTAEVWRQVRDIVERGE 4019
QY 16 ----- 15
Db 4020 QVASFEADVVELGADBSLARLVGSIAMLHGDHQAQAAVGAHLHYVNGVSVBMSAVLGD 4079

```

QY 16 ----- 15
Db 4080 VPTVTRIDLPYAFORHMLEGTDRATAGHPLLGSVRLAASGVLTARYSRSGDM 4139
QY 16 ----- 15
Db 4140 LRQOTVLPATVEMALAADEVGCGIVEDLSVEALLLPDDGAVEQVWGEDEGR 4199
QY 16 ----- 15
Db 4200 RLSTVHARYGDEPTCTATATLATTTGVAANAAGWAGVWPAGAVPGT SAPSLRAV 4259
QY 16 ----- 15
Db 4260 RLGSDFRAVALDADHONTREVLPALMAAALTVEGETRAVNOGLTLHAGNGELAVRL 4319
QY 16 ----- 15
Db 4320 TSHDDSTLSAETDSTGLPVLARSILRLVPEPATSTDDLLTLWAGIPPOQTGLT 4379
QY 16 ----- 15
Db 4380 VGAFEDLAADGDVPEVAVFTALPDSDDPLEOTRKILAOVLTLOELGGERFSDSTLY 4439
QY 16 ----- 15
Db 4440 VRTGGLAAAGVSGLMSAOSGHPGRVJVESDALTQDOLAAAVGLDEPLRVSQRYE 4499
QY 16 ----- 15
Db 4500 VPRLTRHAEEPERETWPDGTVLITGSGVLGIAARHLWBERGVHLLLSRSAPDE 4559
QY 16 ----- 15
Db 4560 ALIGELGELGARVETACDVSDPALQVLAGVSEPHLRAVITAGVVDGVESLTVQ 4619
QY 16 ----- 15
Db 4620 RLETVLRKADGAMNHELTRDADLAFAVMTSAGVLSAGQANVAANAFDLALAEOR 4679
QY 16 ----- 15
Db 4680 RAEGLPALAVANGLEDASGITAOITDTRDIRRGILRAISABHGMLPDSASRSEPV 4739
QY 16 ----- 15
Db 4740 LVAAPEVROAEVPAALLRSILHRPIARBAAGARWLAALAPAREKALLKIVDSAT 4799
QY 16 ----- 15
Db 4800 VLGHADTSTVSVAVFRDLGVDLSLAVELRNSLAKATGLRLPATLYEDPYPTALAVRLG 4859
QY 16 ----- 17
Db 4860 ELFTGENPVPRGVSAAVAODEPLAIVGMACRLPGCVSSPEDMLNLESCTDAVSGEPTD 4919
QY 18 ----- 17
Db 4920 RGMDEVNLYDMAGKSHRAGEGLDAAAFDAGFFGISPREALAMPQORLIVLESWEAFE 4979
QY 18 ----- 17
Db 4980 RAGIEGVSRSGETGVFWGAVPGYGAGADLGFPAATASATSVLSGRVSYFEGLEGPAFT 5039
QY 18 ----- 17
Db 5040 VDTACSSSLVALLHQAGIALROGECSLAIVGYVMATPELFTERSKRGSLASDGRKAF 5099
QY 18 ----- 17
Db 5100 DSADGTGAEGVGLVRLSELSDAQAQKHQVLAVRSASVNDGASNGLTADNGP9QORVI 5159
QY 18 ----- 17

Db 5160 QAALSNAGLAHEVDVNEAHGTGTLGDPLEAQAIVATGQDERPILLGSLKSNIGHAQ 5219
QY 18 ----- 17
Db 5220 AASGVSVIKMVALOHNTVPRTLVDEPSRHDVNAAGAVELRENDPQGTDRPRRAGV 5279
QY 18 ----- 17
Db 5280 SSFGVGTNAHVILIESARPAQPAENQPEVTPVNASDVLPLVISAQTQPALHEHEDRLA 5339
QY 18 ----- 17
Db 5340 YLAASPGADTRAFASTLAVTRSFVFEHRAVLGDDTVTGFAVSDPRVEVFPQGWOMLGM 5399
QY 18 ----- 17
Db 5400 GSALRDSVYFAERNAECALSEFVMDLTVLDPPAVDQVYVQASMAVMSLAAV 5459
QY 18 ----- 17
Db 5460 QAAGVRPAVIGHSGETIAACVAGAVSLRDAARLVTLRQAIRAGLAGKAMASVALPA 5519
QY 18 ----- 17
Db 5520 QDVELVDGAMIAAHNGPASTVLAGTPEAVDHYLTAEARGVRRITVDYASHTPVELI 5579
QY 18 ----- 17
Db 5580 RDELDTSSSQAPLVWMLSTVDGSNVDSPLDGETYRNLREVGTFHPAYGOLQOAGD 5639
QY 18 ----- 17
Db 5640 TVFEVVASPVLLQAMDDVYVATLRRDDGATRALTAQAAYGVTVDPALIGTTT 5699
QY 18 ----- 17
Db 5700 TRVLDPYTAFOHORYWEGVDSAGGHPILGAVELPDSNGVVLGHRVSLATHTWLAD 5759
QY 18 ----- 17
Db 5760 HAVRGVLPGTAFVELVYRAADVECDVIDELVETPLLPOTGVQVLSVAREDSG 5819
QY 18 ----- 17
Db 5820 HRAVTVFSRADNADTWRHVSATISASDAPLSLPERASWPPAQOQPTNGDLYDLAAG 5879
QY 18 ----- 19
Db 5880 TETGPAPQGLQAAMRDGDTNVAEVALAEQOQAFANFVHPALLDAALHASTVLPDPAEQ 5939
QY 20 ----- 23
Db 5940 QSLRMPFSMSHYOVHATGATLKVAMPTTIDGMSVHADDDGRPVATIGSFVTRPYTADA 5999
QY 20 ----- 23
Db 6000 IGSAADLLEVVYTEIPITPOQTGILTGRREDLVADVPVEVVYTYARDTGSSPDPLAQ 6059
QY 24 ----- 23
Db 6060 TRTLTAQVLAQVAMLAGERFTDSTLVKTGTGLAAANSGLMRSVQSEHPGRFTVLEBD 6119
QY 24 ----- 23
Db 6120 DQTLTPQLAATAGLDEPLRVCDSRFEVRLARANTPESSPLTIPDDRAMELEGRSGT 6179
QY 24 ----- 23
Db 6180 LODIALVPTDARPLRPEVRIDVRAAGLNFMDVYLALGTYPGEAVIGAEAGVYLEVG 6239
QY 24 ----- 23

```

Db 6240 PEAHDLAPGDRVGLVGGFGAVALADRMLAVIPDGMSFTTAASVPVVFATAYGLVDL 6299
QY 24 ----- 23
Db 6300 GGLSAGESVLIHAAAGVGMATQIARHLGAQIYATASAGHILYEAGLDGTRIADST 6359
QY 24 ----- 23
Db 6360 TGFREAFNTDGRGVAVVLSLSDGFVDAIDLLPRGGRVEIGKTDIDRPHRTADRP 6419
QY 24 ----- 23
Db 6420 GTTQACDMDVGPDLREIITELISLFGQVLQPLPYQTDIDQARDASMSRAHIG 6479
QY 24 ----- 23
Db 6480 KIVLTIPRRPDDTILITGSGVLAGILARHLAEHGARHLLLSRTAPDEALIKELAE 6539
QY 24 ----- 23
Db 6540 LGARVETACDVSDRAGIARVAGVSPHPLTAVIHTAGALDDGVESLITQQLDVLRLP 6599
QY 24 ----- 23
Db 6600 KADGAMHIELTRADLAFAVYSSAAAVLGNBGGNYAANAFULDALAQRTQGLPAL 6659
QY 24 ----- 23
Db 6660 ALAMGPMEYTDLTQALNGTDQDRIRCSGKRTTAEIDGNRLFDTASHGPELTPAVLDP 6719
QY 24 ----- 23
Db 6720 TRDGEVPLALSLRRPIARRASADGGVQMLALAPAREKALKIVCSAAWVLGHADA 6779
QY 24 ----- 23
Db 6780 RSIPAGAFKDLGVDSLMAYELRNGLVKATGLRLPATLVYDTPPYLAARLDELFTGEN 6839
QY 24 ----- 23
Db 6840 PAPVRGVSVDGDEPLATVGMACLRPGVSSPDLKRLVESGTDAVSGFTDNGMDVEN 6899
QY 24 ----- 23
Db 6900 LYDSDPEAGSYCYGGFIDTAGFDAGFFGISPRALANDPQKLLLEVSNEAFERAG 6959
QY 24 ----- 23
Db 6960 IEPGSVKSDDTGVFAGFVGYGAGFDREGATSGPSVLSGRVSYVFLGSPALIMDTA 7019
QY 24 ----- 23
Db 7020 CSSSLVALHLAQAALRNGECSSMALAGVTVMAPEVETEFARORGLASDGRCKAFADSAD 7079
QY 24 ----- 23
Db 7080 GAGFSEGAQLLVTERLSDARRNHQVLAVVRGSAVNDGASNGLTAPNGSQRVIRBAL 7139
QY 24 ----- 23
Db 7140 SNAGLSTADVDVEAHGTGTLIADPIEAQALLATYGGDRBQPLLGLSKNIGHTQAASG 7199
QY 24 ----- 23
Db 7200 VSGVTKVMYALRIGFVFRILHDEPSRHNDMAAGAVELRENQPMGTDPRRAGVSSFG 7259
QY 24 ----- 23
Db 7260 VSGTNAHVLESAPPAQPAEEQPVETPVVASDVLPLVISAKTPALTEHEDLRATLAA 7319
QY 24 ----- 23
Db 7320 SPGADTRAVASTLAVRSVVERRAVLLGDAVVTGTAVTDPRVVFEVPGGGMQMLGMSAL 7379

```

```

QY 24 ----- 23
Db 7380 RDSVYFAERMAECALASEFTMDLFAVLDPAVDVDRVDVQPASAAVWSLAAWQAA 7439
QY 24 GIRD----- 27
Db 7440 GVRPDVAVIGHSGEIAAACVAGAVSLMDARITLRSQATIRGLAGRAAASVALPAHEI 7499
QY 28 ----- 27
Db 7500 ELDGAMIAAHNGPASTVAGTPEANDVFLTAHEANGVRVRITVDYASHTPVELIRDE 7559
QY 28 ----- 27
Db 7560 ILGTTAGISGPPVYVRLSTVDSWVDSPLDGEYWTNRLNEPVGFHPAVSQAQGDAAVF 7619
QY 28 ----- 27
Db 7620 VEVASPVLLQAMDDVYVVAFLRBDGDATRLTLAQAVYHGVTVDMPALIGTTANV 7679
QY 28 ----- 27
Db 7680 LDLPYAFQHQRYVAVSVDRAADGPHLLGAAYELPESDGLITGRVSLATBAKLADHAV 7739
QY 28 ----- 27
Db 7740 WGRVLLPGRFVFLVYHAGEVCCDVYDELVTETPLLPTQGVGLSVSGENDESHRV 7799
QY 28 ----- 27
Db 7800 VTFVRADNADTWLRHVSATVRVSDTTPPSDLTAMPRAQRPVDAAGFYDQLTGMYEV 7859
QY 28 ----- 27
Db 7860 GRPFGQLQAAWNGDITVEAEVALAEQVREAAVAVHPALDLAALHACTLNASDAEVGVG 7919
QY 28 ----- 27
Db 7920 LPEFSGVVRHAGSAMLNVAVTQAADGMSVYVADDIGRPVAVSGSLVTRPTADNLGSA 7979
QY 28 ----- 31
Db 7980 ADDLLALTAGIPTPQOTGLTWGRF 8004

```

RESULT 13
 T20802
 hypothetical protein F12F6.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20802
 R:Kershaw, J.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: 219326
 A:Accession: T20802
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1059 <WILL>
 A:Cross-references: EMBL:273425; PDB:1CMA97790.1; GSPDB:GN00022; CESP:F12F6.5
 A:Experimental source: clone F12F6
 C:Genetics:
 A:Gene: CESP:F12F6.5
 A:Map position: 4
 A:Introns: 57/1; 437/3; 741/2; 984/2; 1022/2

Query Match 39.1%; Score 61; DB 2; Length 1059;
 Best Local Similarity 1.9%; Pred. No. 3.7e+02;
 Matches 14; Conservative 8; Mismatches 8; Indels 699; Gaps 4;

```

QY 1 SRX----- 4
Db 234 SRHRSLVYKTFRRREKEFVYVRLCTKARNETLLCVKANAALHRRFAQDLSTLIDCML 293

```

```

QY 5 -----XHSME----- 9
Db 294 GMDPMLPALLEKVIDERKKTITTOHEMDSLASTILRSSVDYKADKOFPEANHQLFMLPKQ 353
QY 10 ----- 9
Db 354 FERRPOLGDDIMEVSADQSISDQLQKQIQIEKRLBGLQFEVDEYVKKSLSEKQLLQIX 413
QY 10 ----- 9
Db 414 NTQPDGKGRNDLHYTYOYLKFFENFLNGMRLPARSASIGELGKNGISLKSQA 473
QY 10 -----XRTPD----- 14
Db 474 SVGSASERNGSTTENGSSQNHLLERRSKARIGIYVKSPPDRPRPLFGGSIDDEV 533
QY 15 ----- 14
Db 534 EATGEIPLVQSAIAYISRYSLNQGLFRVSGSQSEINFRAYERGEDLFQYLDGSD 593
QY 15 ----- 14
Db 594 ANSAGVKKIYFRLEBPPEPFPEQFCDCANSESPTEFRARRELVSKLPSHVILLR 653
QY 15 ----- 14
Db 654 FLFAPLSLCEADENKMPHNLACFPPLLP IPEGKQVYVANYVELVRNLIHADD 713
QY 15 ----- 14
Db 714 VEPRLAGPYDKYAMQRYMDGNFLENDLISEDEAHEKLSLPPSRHMIDSTYESADR 773
QY 15 ----- 14
Db 774 ILLSPILLSQANTSPNCSISNCAESSTNNEIDYAPLASRSSNRKASGSDILMARSE 833
QY 15 -----INPATYXKR----- 23
Db 834 MPRIANELNMFKNSSLSDESGKISVLRNSHVEPSTWDRKHLNKRMSSTDPEREY 893
QY 24 ----- 23
Db 894 VSPPEPLISTRYVSAASCISPAFPTIDIKERKDSRLLVAPITRATVPANGSNTEN 953
QY 24 --GIRPGR 30
|:|:|:|:|
Db 954 STGVPRICK 962

```

RESULT 14

B96695

hypothetical protein F5A8.4 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence: revision 02-Mar-2001 #text: change 31-Mar-2001

C:Accession: B96695

R:Thelogy: A.; Ecker, J.R.; Palm, C.J.; Pedersen, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.A.

ansen, N.F.; Hughes, B.; Nulzar, L.

Nature 408, 816-820, 2000

A:Authors: Number, J.L.; Jenkins, J.; Johnson-Nopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lirios, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, N.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Ventier, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: B96695

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-5138 <STO>

A:Cross-references: GB:AE005173; NID:g4204276; PID:AD10657.1; GSPDB:GN00141

C:Genetics:

```

A:Gene: F5A8.4
A:Map position: 1
Query Match 39.1%, Score 61; DB 2; Length 5138;
Best Local Similarity 0.7%; Pred.No.7.5e+03;
Matches 17; Conservative 4; Mismatches 10; Indels 2303; Gaps 4;
QY 1 SRKH----- 5
|:|
Db 1439 NPLHGRITLVARDLSWAFVNMATESIGPAVALHGAFLVLDGLSLGTGFSGRDGL 1498
QY 6 ----- 5
Db 1499 RKKCAFLLQLELPASDTPLELSRMELYGWDSKAICRSKSVRBMFGIDPFITSK 1558
QY 6 ----- 5
Db 1559 GDNPEIGFEFLAFTTHRVNLAVLRAMQLSKPILLESFGVGTSLALGKXSGHKV 1618
QY 6 ----- 5
Db 1619 RINLSQTMMDLQSDLPVSEDEMKFAMSDGILLQGLNALIDHRAQVFIPELGTEEC 1678
QY 6 ----- 5
Db 1679 PPTFRVACQNPSTQGGKGLPKSFLNFTKVYVDELVEDDYLFICSLYPSVPSILS 1738
QY 6 -----HSM----- 9
|:|
Db 1739 KILALNQLDGMVLRKNGHDSPEWFLRDVIRSCQMOEATIDLFESFLAVLYIQ 1798
QY 10 -----XRTPDIN----- 17
|:|:|:|
Db 1799 MRYATRKKEVLRIRYKAFIDKTPSINPYRQVLPATLVGTALIKRNLSNIAEQEL 1858
QY 18 ----- 17
Db 1859 LPEIRONLEAVACQONKLCIIVGSSSGKTSVIRILAQLTGYPLNELNSATDSDL 1918
QY 18 ----- 17
Db 1919 LGCFEGYNAFRNRYMTRVENLVDEYNSLLQSSQELFSNRGLVSWLSYLNKIDS 1978
QY 18 ----- 17
Db 1979 LVNPLFLNDSSTLSEVEYEDLPQVLEGLVPYSKXYLEQISTITLQTHKKQ 2038
QY 18 ----- 17
Db 2039 STFEWVTGMLAIKKGEMVYLKNNLCNPVYLDRINSLVPCGSITINEGYNPEV 2098
QY 18 ----- 17
Db 2099 TVVPHNRFLFLVNPKEFVSRAHNRNGVEVPMQPHQVNDGSCBELVIGVEREL 2158
QY 18 ----- 17
Db 2159 ALSGIDGYKLVTSMAAAVHAWLNGOSFGVRYTYLLEQVHLFQULLNNGQLMSLQL 2218
QY 18 ----- 17
Db 2219 SWEHYILSLGVTDKGVYDFVETVLSDELSDSTMGDLYLPGKPRPNLRLTW 2278
QY 18 ----- 17
Db 2279 YSRRTYRQNCMTLEFAGQYASHQKISDNVKSRELAAGEPRRIYSIDSITLKKVLF 2338
QY 18 ----- 17
Db 2339 PKALIGSSCAPDAANFENDLASMLLPANNTTEQATEEDIQLYLAFSFGSRIQOHP 2398
QY 18 ----- 17

```



```

Db 2399 FLCCFLNTLKVEFENPIMNHISRCRKNLFCRLDDPAVPIPMLSKLLIDVAASNDQSKP 2458
QY 18 ----- 17
Db 2459 YKSLFESLNSVGLNRSTQWLVESNDNHVDVSTFRRLDRLVLEKLLICEIYGASF 2518
QY 18 ----- 17
Db 2519 SVLIQITVEIINDNSFWSGLVSSDEYLLFEFWSLINSIKKMSFFPGEVQVLEESKN 2578
QY 18 ----- 17
Db 2579 INNVIHGHPEKSMVMAYGHSPLFVSALFPHKQOEFLQICSTVWPLKSESDENGNDHIT 2638
QY 18 ----- 17
Db 2639 KAIPFSGELCLALEGICISSYINDEDDVDVAVAQDEIYQFLERLLEKRLKEDKM 2698
QY 18 ----- 17
Db 2699 GFSEIDNTENTACCCVCPETVTTGSGFSWVKTCFTASSESGSLDYELLAALQHLIVA 2758
QY 18 ----- 17
Db 2759 RPTHEQDLVDIRKLKLPALVEYSLSTRPQTLVAHQKLLMAIDAHSELGVDTIAGPAL 2818
QY 18 ----- 17
Db 2819 EIKTWHSVLKNSQIGIMIAHLVTSFELMTGPTETLRHMQLSDSVTFYADSACENSY 2878
QY 18 ----- 17
Db 2879 RCDSDIITHQSFESSEFVAIKSVFHALEKKQKMDGIONLISLIGSSHNKIKSVTHS 2938
QY 18 ----- 26
      ATYXKRGIR
Db 2939 FVGLPLAKRLYSSSSNFEYCNGLAMLYLGLRPHLNSLDVIDPAMKICCKILKEKI 2998
QY 27 ----- 26
Db 2999 SSELNFKVRGECYLSGLYSNNDSSEHTLSKLTERRKLRQKVIYFNSDPKQYDLR 3058
QY 27 ----- 26
Db 3059 RALDEFAGELFRPISLVNDIEVLDMQVVEQVFMQETALSIFIDRMSDSEYVDITQPI 3118
QY 27 ----- 26
Db 3119 QVSYTEMKLGLSLFVSGALLGKLLNRPDIDWDSVMEIYALNRPDSSIASFTYTECL 3178
QY 27 ----- 26
Db 3179 PPLHLISGANSRAKSLGDLVGLLKLISVSSAEDSRKASLOKVALYKMLHAKVYLQFVA 3238
QY 27 ----- 26
Db 3239 NTGLLDEASPELLDKIYVELARIMENKFOAKTKADNLPFLYKFRSDFKIDSVMEDIS 3298
QY 27 ----- 26
Db 3299 ALGKTFPNSFSSEWQELADDDIKNVKMDTHIDDEENLEDMDLLIOEHDSIYSTHNEL 3358
QY 27 ----- 26
Db 3359 FGFCDLSEKTIYISQSGRCITDSRLDSTFDSYELGVSMIKRLGLFTSSLDKLVPEH 3418
QY 27 ----- 26
Db 3419 LRLCLENKKNFTSNYSQASKYNFYKDLDPGLGKMKVFLTPIQORINSLLQEREDHPGL 3478
QY 27 ----- 26
Db 3479 QKLSGVQLMLALPSSITPLAKALSGLOFLCKVYKLOEGCKLPISDLLEPTIISLASSMQ 3538

```

```

QY 27 ----- 26
Db 3539 KVEERKRPILLDEVDQOYELNARKFCRRMLMKPQNKISPFHKVSFLDKGLQHYIILK 3598
QY 27 ----- 26
Db 3599 SSQLIAPLEFLQILSNLTFPDSRFYFDSKHKRVENYINFGYIOFLPVNEQDLNR 3658
QY 27 ----- 26
Db 3659 KNVETELKEVLKICRMRBPNLYNETTKTRQKVKLLQKFTDMRLPVLVYKPDUTRE 3718
QY 27 ----- 31
      PYGRF
Db 3719 RAQFLPILLDPLDMDGASDMRIEVLVSALDAEQLRDRSSRYVWMNKLKESVGR 3772
      IIII

RESULT 15
A:88852
protein unc-22 (imported) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A88852
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000: NCBI:99069613, PMID:9851916
A:Note: see websites genome.vustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: A88852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: GB:chr_IV; P1DN:CAA98081.1; P1D:g3881830; GSPDB:GN00022
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homolog

Query Match          39.1%; Score 61; DB 2; Length 6831;
Best local similarity 0.3%; Pred. No. 1.3e+04;
Matches 16; Conservative 6; Mismatches 9; Indels 4954; Gaps 4;

QY 1 SRXHXSMR----- 9
      |||
Db 1127 SRNGTHLKIKLENNEDVYEIDQGLGSCNVTYLEAKRPILNWKPKTEAKAGEPC 1186
QY 10 ----- 9
Db 1187 VVKVPQIKGRGDKPAQILNKGKPIDEMKRIEVIKKDVAIEVFNQPLADTGKVA 1246
QY 10 ----- 9
Db 1247 LEIGNSAGTALAPFELPVADKRPKPGPLEYKNVTAEGIDLWGTDPDDEGAPVKAAYIE 1306
QY 10 ----- 9
Db 1307 MOGSGNMAKVGETGTFKVKDLKEHGEYKFRVKALENGSLDPLTGESYIAKNPYGV 1366
QY 10 ----- 9
Db 1367 PGKPKMDAIDVDKDHCTLAWBPREDGAPITGYIERREKSEMDWQVQITKPDCEL 1426
QY 10 ----- 9
Db 1427 TDKVVEDEKYLIVKAVNKAQPDGCDHGKPIKAKAKASPEFTGGIKDLRLKVGFTI 1486
QY 10 ----- 9
Db 1487 KYDVPISGEPLPECLMVVNGKPLKAVGKVKMSSEGHKHKIENAVRADSGNFTITLKS 1546
QY 10 ----- 9

```

Db 1547 SGSCSTATVTVGRPRPKPLIDADVACDAGATLSMNPDDGGDLTGVIYEAQMDN 1606
QY 10 ----- 9
Db 1607 KGYIENGKVDPRNTTLKVNLKRNKYNKRVKAVNNEGESEPLSADQYQIDPMDPG 1666
QY 10 -----XRTPDIN----- 16
Db 1667 KPGREITDFDADRIDIANEPHKGOGAPIEEYIYEVDPDTKMKREKVPDTNASSIG 1726
QY 17 -----PAW----- 19
Db 1727 LKEGEQFRVAVNKAQPGQSPSEKQIAKFIIPALIKHDLKSIYKAGATVEREV 1786
QY 20 ----- 19
Db 1787 KIGEPPIEVAKFGNOQLENGJQLTIDTRKNEHTILGIPSANSDVGEYRLTYKNSHA 1846
QY 20 ----- 19
Db 1847 DEKAMLTVDLRPSKPNPLEVSDVEEDNMLSKPPDDGGEPJEYIYEVKLTATGRW 1906
QY 20 ----- 19
Db 1907 VPCAKVDTKAHIDGLKKGQYQFRVAVNREGASDALSTDKTRAKNPYDEPGTGPD 1966
QY 20 ----- 19
Db 1967 VYDMDADRSLJEWEPKSDGAPITQYVIEKKHGRDMQCGKVSIGDQTNALILGLKEG 2026
QY 20 ----- 19
Db 2027 EBYQFRVAVNKAQGEASDSKAVAKPMLKPWIDREAKTITIKYGNVDEVDYVRG 2086
QY 20 ----- 19
Db 2087 BEPRKEMIFNEKPYVDOKIRIESEYKIRFVLNGATRKHAGLYTLATNASSGDSKESVE 2146
QY 20 ----- 19
Db 2147 VVILCKPSPILGPLEVSNVIEDRADLEMKVPEDDGAIDHIEIKMLATGRWPCGRS 2206
QY 20 ----- 19
Db 2207 ETTTTPMLQPGHEKFRVAVNKEGSDPLTNTALIAKNPYEVPGHVDPBELVMDK 2266
QY 20 ----- 19
Db 2267 DHVLAWNAAPDDGAPIEAFVIEKKDKNGRMEALVPGDKTATVPMLKRCGEYQFRIS 2326
QY 20 ----- 19
Db 2327 AKRKAGTGPSPDSRVAKPBNLAPRIHREDLSDTYKVALILKEIYHIDGEPAPDVTW 2386
QY 20 ----- 19
Db 2387 SFNGKIGESKAQIENEPIYSRPAIPKALKQSGKTTATNINGSVTTINIKVSKPT 2446
QY 20 ----- 19
Db 2447 KKGPIEYTVDEDRATLDMKPEDDGGEPJEFYIEIKMKNKDGIVNFCGNSGJTHFTVD 2506
QY 20 ----- 19
Db 2507 SLKNGDHYFRKAVNSESPDLFETDILAKNPFDRDRGRGEPDMDSDHVDLMD 2566
QY 20 ----- 19
Db 2567 PPLSDGAPITEYOIEKTRKGRMEBALTVPQOITATVPDLTPNEEYFRVAVNKAQ 2626
QY 20 ----- 19

Db 2627 SDPSASAKAVIAKPRNLKPHIDRALKMLTIRAGOSISFDVVPSEPAVTWMPDRE 2686
QY 20 ----- 19
Db 2687 IIRNGRVKIDNPEYQSKLVVKOMERGDSGTITIKAVNNGEDEATKINVIDKPISPNGP 2746
QY 20 ----- 19
Db 2747 LDVSDVHGDVTLNWRAPDDGGIPIENYVIEKYDASGRWPAKAVAGDKTAVVDLI 2806
QY 20 ----- 19
Db 2807 PGHEYFRVAVNAGESEDPLETFGTTLAKDPDKPGKTINABEIIDMDKDHVLEMKRPA 2866
QY 20 ----- 19
Db 2867 NDCGAPIEEYVEMKDEFSPFNNDVAHVPAQOTNATVGNLKGSKYEPRIRAKNAGLGD 2926
QY 20 ----- 19
Db 2927 PSDSASAVAKARNVPVIDRNSIOEIKVAGODFSLNIPVSGEPPTITWTEEGTPVESD 2986
QY 20 ----- 19
Db 2987 DRKTLNEDGKTYHVKRALRSDTGTIICAENENGTIAEVKATVLDHPSSRGRPLDVT 3046
QY 20 ----- 19
Db 3047 NIVKDGDLAWKPEPDGGEISHYVIEKQDAIGRTACGSKDINFVDDLTQGEYK 3106
QY 20 ----- 19
Db 3107 FRVAVNRHSDSPLERKALIAKDFDRADKGPPEIYDMDKHADLKTTPPADGGAP 3166
QY 20 ----- 19
Db 3167 IEGYVEMKTPSGDWYPAVTGAGELIAYVDGLKPGQYQFRVAKLNKAGESTPSPSRT 3226
QY 20 ----- 19
Db 3227 MVAKPRHLAPKINRDMFVAGRVAGOTLNFVNVGEPAKLEWFLNGSPLSSGNTIID 3286
QY 20 ----- 19
Db 3287 NNTDNTKILTKSTARADSGKYIATNBSGKDEHVDNIIIDIPAGPGLRHNDITKE 3346
QY 20 ----- 19
Db 3347 SVYLAKDEPLDDGSPITNTYVEKQEDGGRWPCGETSUTSLKVNLSGHEKFRVAV 3406
QY 20 ----- 19
Db 3407 NROGTSAPLTSDBAIVAKNPFDEPDAPTVTPVDMDKDHVLEMKRPANDGAPIDATYIV 3466
QY 20 ----- 19
Db 3467 EKKDEGDWYECARVDSKTTKATADNLTPGETYQFRVAVNKAQPGKPSDPTGNVYAKR 3526
QY 20 ----- 19
Db 3527 RMAPKLNLAGILDIRIKAGTPIKILDLAFEGEPAPVAKWKANDATIDGARADVTNPTSS 3586
QY 20 ----- 19
Db 3587 AIHIFSAVRGDTGVYKILIVENEGKDTAQCNAVYLDVPGTPEGPLKIDILHKGCTLNNK 3646
QY 20 ----- 19
Db 3647 PPTDNGETDVLHTYVEKADTSRGTWDEVGTFPDCIAKVNKLVPCKEYAFRVAVNNGES 3706
QY 20 ----- 19
Db 3707 KPLAEERPIAKNOFDVDPVDKREVTDMDKRIDIMNPIANNNGAPVTGYIYEKKEG 3766

```

QY 20 ----- 19
Db 3767 SAIVTEAGKIPGTTFSADNKLKPYEYERVAIVANAAGSPDSDFDQITKARTLKPKIL 3826
QY 20 ----- 19
Db 3827 TASRIKIKAGFTNNLEVDFIGADPTATWVGSGAALAPELLVDAKSSSTISIFPPSAK 3886
QY 20 ----- 19
Db 3887 RADSGNTKLKYNLEGEDEALFEVIVODRPSAPGPLEVSDYTKDSCVLMWPKPDGGA 3946
QY 20 ----- 19
Db 3947 EISWYVEKRDKTNTWVPVSAFVGTSTVPKLTGSEYEFVMAENTGSRDSLNTDE 4006
QY 20 ----- 19
Db 4007 PVLAKDFGTGPKRPEIYDJDNDHIDIKMDPPRONGSPVDNYDIERDKATGRITKY 4066
QY 20 ----- 19
Db 4067 NTSVQGTASDTRVQGHTEYERVAIVANAAGSPDSASAATAKPMHAPKFDLDLG 4126
QY 20 ----- 19
Db 4127 KEFRVAGEPLVITITPTASPODISWTKSGKPLAGVETTSQTKLVIPSTRSDGPRV 4186
QY 20 ----- 19
Db 4187 KIKAVNYGEAEANIKITVYIDKPGAPENITYPAVSRHTCTLMNDAPKRDGAGIAGYKIE 4246
QY 20 ----- 25
Db 4247 YOEVSQIMDKPGLISGATVYRGLSEHQYFRIRAEVAGLSDYCOGVPIYKDPED 4306
QY 26 ----- 25
Db 4307 PPGAPSPETITGYDINQVSLAMNPPRDGSGPLIGYVERFEKRGGDMAVPMVYKGT 4366
QY 26 ----- 25
Db 4367 ECIYVGLHENETQYFRVAVANAAGSPDSBPVTCRPYERKAPADPRVGIKTKNSA 4426
QY 26 ----- 25
Db 4427 ELTWNRPLRDGAPLDIGYIYEKKLGDNDMTKCNKPRVDTAFEVKNLGEKEFERVIA 4486
QY 26 ----- 25
Db 4487 VNSAGEBPSPSDVLIBQPGRPITFDINNLKIDIVRAGETIQIKIYAGNPKPIIDL 4546
QY 26 ----- 25
Db 4547 FENSPIFENERTVDVNGEIVITTTGSKRSDAGYKISATNKYKDYCKLWVFLDAP 4606
QY 26 ----- 25
Db 4607 GKFTGIRATDQADAMTLSWPPKPDNGDALTINVEKRTGEGDVIYVGHVGTTLVRV 4666
QY 26 ----- 25
Db 4667 NLDANTYERVRANQYGVGEPLFETDIAIVAKNPFDTGAPGQPEAVETSEALITLQWT 4726
QY 26 ----- 25
Db 4727 RPTSDGAPITQGVIEKREYGESEWTKAFAFGNLDTKNRYGLTPTKTYEFVAVANAG 4786
QY 26 ----- 25
Db 4787 QGEYSVNSVPTADNAPTRPKIMGMLTRDILAYAGERAKIIVPPAASAPVVTTSKGEN 4846

```

```

QY 26 ----- 25
Db 4847 KISPTDPRVEXSDEIATLTIKESLTDGLYFVELNSGSDSASIRLKYVDKPARSQ 4906
QY 26 ----- 25
Db 4907 HIRVEDIADCCCTLYMPPSSDGGSPITNYIVEKIDLRHSDGKWEKVSFVRNLNVTYVG 4966
QY 26 ----- 25
Db 4967 LIKDNRYFRVRAEYQYVSEPCELADVVYAKYQEVNPQAPVYRDKDSTWAELENDP 5026
QY 26 ----- 25
Db 5027 PRDGSKIIGYOYVNDTSSGRWJNAKMDLSEQCHARVTLGLQNGEPEFRJIAKNAAGPS 5086
QY 26 ----- 25
Db 5087 KSPSPERCQLKRGPPGPIHWGAKSIGRNCTTWMAPLEDGSKITGYNVEIREYG 5146
QY 26 ----- 25
Db 5147 STLMTVASUYNREPEITVDKLBREFNDYFRVYALNAGKIGTSLPSGFIKIQESGSRP 5206
QY 26 ----- 25
Db 5207 QIVVYREDTQAPYNNRAVTCBAVGRPEPTAKMLRNGRELPESSRYRFEASDGVYKFTIK 5266
QY 26 ----- 25
Db 5267 EYWDIDAGIYEVSNFYSGDTATNLVYQAPVYIEKQVPTILPSGDLVRLKIYFSGTA 5326
QY 26 ----- 25
Db 5327 PERHSVLNREELMDHPTIRIVERFDHLITIPALSVREAGRYEYVNSDGEATGFW 5386
QY 26 ----- 25
Db 5387 LNVGLPEAPQGLHISNIGBSTALSWRPVYTDGSKITSYVEKRLDSKDEWTVTSN 5446
QY 26 ----- 25
Db 5447 VKDNVYITGLFENHEFERVSAQENEGICAPVSENPIARLPDPPTSPLEITQVG 5506
QY 26 ----- 25
Db 5507 GDYVTLNQRPLSDGGRLRGYIVEKQEBHEDEWFRCONQSPNNVNPVPLIDGRKRY 5566
QY 26 ----- 25
Db 5567 KVFVANDAGLSDLAELDQTLQASGSGEPRKIVSPLSDINEVGRVYFECESISPRPE 5626
QY 26 ----- 25
Db 5627 YWFKGCELVDTSKYTLINKDKOVLIINDLISDADDETGRAFNSSGTRSTRANLRK 5686
QY 26 ----- 25
Db 5687 TKPRVITPKYHGYEAKGETIELKIPYKAYPOGEARWIKGEKIEKNSKFSITTDKF 5746
QY 26 ----- 25
Db 5747 ATRISNASREDYGERVVENSVSGDSGTIVNADVPEPPRPPIENILDEAVILSWR 5806
QY 26 ----- 25
Db 5807 PPALDGSIVNTYTIKREAGSGWSPCAKSRYYTITBGLRAGQYEFRIANKNQGS 5866
QY 26 ----- 25
Db 5867 KPCEPTAVLIPGDERKRRRGYDVEQGIYVKGQVSSNIDNVFDIMQYQYPOVPEIK 5926
QY 26 ----- 25

```

```

Db 5927 HDHVLDDHYDINHEHGTGANGVVRHTERATGNNFAAKFVMTPHESDKETVRKEIOFMSVL 5986
QY 26 ----- 25
Db 5987 RHPTLVNLDHAFEDNDENMVIYEMSGSELEKVADEHNKMSDEAVETMRQVCKGLCHM 6046
QY 26 -----R 26
Db 6047 HENNVYHLDLKPENINFTTKRSNELKLIDFGLTAMLDPROSVKVTGTAEFAAPEVAEGK 6106
QY 27 PVGRF 31
   III
Db 6107 PVGYT 6111

```

Search completed: May 1, 2003, 14:49:53
 Job time : 69 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using SW model

Run on: May 1, 2003, 14:49:32 ; Search time 45 seconds

(without alignments)
63.277 Million cell updates/sec

Title: SEQ74-PLUS-73

Perfect score: 156
Sequence: 1 SRHXHSMKRPDINPAMYXKRGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.0

Searched, 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications JA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	31	10	US-09-932-161-14
2	147	94.2	31	10	US-09-932-161-15
3	147	94.2	31	12	US-10-044-592-4
4	147	94.2	31	12	US-10-044-592-5
5	147	94.2	70	12	US-10-044-592-90
6	147	94.2	82	12	US-10-044-592-96
7	147	94.2	86	12	US-10-044-592-96
8	147	94.2	87	12	US-10-044-592-92
9	147	94.2	91	12	US-10-044-592-92
10	146	93.6	31	10	US-09-932-161-13
11	146	93.6	31	12	US-10-044-592-39
12	146	93.6	32	12	US-10-044-592-40
13	146	93.6	33	12	US-10-044-592-41
14	146	93.6	98	12	US-10-044-592-28
15	146	93.6	98	12	US-10-044-592-38
16	146	93.6	98	12	US-10-044-592-82
17	146	93.6	98	12	US-10-044-592-84
18	146	93.6	98	12	US-10-044-592-86
19	146	93.6	98	12	US-10-044-592-88

20	135	86.5	29	12	US-10-044-592-26	Sequence 26, Appl
21	116	74.4	25	12	US-10-044-592-78	Sequence 78, Appl
22	105	67.3	20	10	US-09-932-161-18	Sequence 18, Appl
23	104	66.7	20	10	US-09-932-161-16	Sequence 16, Appl
24	104	66.7	20	10	US-09-932-161-17	Sequence 17, Appl
25	104	66.7	20	12	US-10-044-592-6	Sequence 6, Appl
26	104	66.7	20	12	US-10-044-592-42	Sequence 42, Appl
27	104	66.7	21	12	US-10-044-592-43	Sequence 43, Appl
28	104	66.7	22	12	US-10-044-592-44	Sequence 44, Appl
29	98	62.8	19	12	US-10-044-592-27	Sequence 27, Appl
30	81	51.9	40	12	US-10-044-592-80	Sequence 80, Appl
31	68	43.6	11877	10	US-09-860-846-6	Sequence 6, Appl
32	68	43.6	11877	10	US-09-861-289-6	Sequence 6, Appl
33	68	43.6	12199	9	US-09-988-384B-6	Sequence 6, Appl
34	66	42.3	870	9	US-10-097-534-12	Sequence 12, Appl
35	66	42.3	906	9	US-10-185-050-48	Sequence 48, Appl
36	66	42.3	26926	9	US-09-759-508B-2	Sequence 2, Appl
37	64	41.0	7968	9	US-10-077-130-5	Sequence 5, Appl
38	62	39.7	2969	9	US-09-738-626-4434	Sequence 4434, Ap
39	61	39.1	7257	9	US-10-014-717-5	Sequence 5, Appl
40	60	38.5	1410	9	US-10-014-717-3	Sequence 3, Appl
41	59	37.8	371	9	US-09-966-782A-15	Sequence 15, Appl
42	59	37.8	371	9	US-10-212-980-8	Sequence 8, Appl
43	59	37.8	372	9	US-09-966-782A-16	Sequence 16, Appl
44	59	37.8	372	9	US-10-212-980-9	Sequence 9, Appl
45	59	37.8	940	10	US-09-815-242-10447	Sequence 10447, A

ALIGNMENTS

```

RESULT 1
US-09-932-161-14
; Sequence 14, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:
; APPLICANT: Clivell, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods For
; TITLE OF INVENTION: Promoting Wakefulness and Sleep
; FILE REFERENCE: P-DC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; PRIOR APPLICATION NUMBER: 2001-08-17
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Rattus
US-09-932-161-14

Query Match          94.2%  Score 147,  DB 10,  Length 31:
Best Local Similarity 83.9%  Pred. No. 9.4e-11,
Matches 26, Conservative 0, Mismatches 5, Indels 0, Gaps 0,

OY 1 SRHXHSMKRPDINPAMYXKRGIRPVGRF 31
Db 1 SRAHQHSMKRPDINPAMYXKRGIRPVGRF 31

RESULT 2
US-09-932-161-15
; Sequence 15, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:
; APPLICANT: Clivell, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods For
; TITLE OF INVENTION: Promoting Wakefulness and Sleep
; FILE REFERENCE: P-DC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161

```

;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO: 15
;; LENGTH: 31
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-932-161-15

Query Match 94.2%; Score 147; DB 10; Length 31;
Best Local Similarity 83.9%; Pred. No. 9,4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTDPDINPAMYXXRGIRPVGRF 31
Db 1 SRAHQSHMETRTDPDINPAMYXRGIRPVGRF 31

RESULT 3
US-10-044-592-4
; Sequence 4, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Murine
US-10-044-592-4

Query Match 94.2%; Score 147; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 9,4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTDPDINPAMYXXRGIRPVGRF 31
Db 1 SRAHQSHMETRTDPDINPAMYXRGIRPVGRF 31

RESULT 4
US-10-044-592-5
; Sequence 5, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:

;; SEQ ID NO 5
;; LENGTH: 31
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; NAME/KEY: PEPTIDE
;; LOCATION: (1)..(31)
;; OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 94.2%; Score 147; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 9,4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTDPDINPAMYXXRGIRPVGRF 31
Db 1 SRAHQSHMETRTDPDINPAMYXRGIRPVGRF 31

RESULT 5
US-10-044-592-90
; Sequence 90, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 90
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Rat
US-10-044-592-90

Query Match 94.2%; Score 147; DB 12; Length 70;
Best Local Similarity 83.9%; Pred. No. 3e-10; 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTDPDINPAMYXXRGIRPVGRF 31
Db 22 SRAHQSHMETRTDPDINPAMYXRGIRPVGRF 52

RESULT 6
US-10-044-592-1
; Sequence 1, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 1

LENGTH: 82
TYPE: PRT
ORGANISM: Murine
US-10-044-592-1

Query Match 94.2%; Score 147; DB 12; Length 82;
Best Local Similarity 83.9%; Pred. No. 3.8e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMTXRGIRPVGRF 31
DB 21 SRAHSHMETRTPDINPAMTXRGIRPVGRF 51

RESULT 7

US-10-044-592-96
Sequence 96, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 96
LENGTH: 86
TYPE: PRT
ORGANISM: mammalian
FEATURE:
NAME/KEY: misc_feature
LOCATION: (124)..(1243)
OTHER INFORMATION: Bracket region depicted in FIG 39.
US-10-044-592-96

Query Match 94.2%; Score 147; DB 12; Length 86;
Best Local Similarity 83.9%; Pred. No. 4.1e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMTXRGIRPVGRF 31
DB 22 SRAHSHMETRTPDINPAMTXRGIRPVGRF 52

RESULT 8

US-10-044-592-92
Sequence 92, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 92
LENGTH: 87

TYPE: PRT
ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 94.2%; Score 147; DB 12; Length 87;
Best Local Similarity 83.9%; Pred. No. 4.1e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMTXRGIRPVGRF 31
DB 23 SRAHSHMETRTPDINPAMTXRGIRPVGRF 53

RESULT 9

US-10-044-592-94
Sequence 94, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 94
LENGTH: 91
TYPE: PRT
ORGANISM: mammalian
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(31)
OTHER INFORMATION: primer
NAME/KEY: misc_feature
LOCATION: (925)..(955)
OTHER INFORMATION: Insert fragment of pmg3 as depicted in FIG 36 and 37.
US-10-044-592-94

Query Match 94.2%; Score 147; DB 12; Length 91;
Best Local Similarity 83.9%; Pred. No. 4.4e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMTXRGIRPVGRF 31
DB 22 SRAHSHMETRTPDINPAMTXRGIRPVGRF 52

RESULT 10

US-09-932-161-13
Sequence 13, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Clevell, Olivier
TITLE OF INVENTION: Screening and Therapeutic Methods for
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0

Thu May 1 15:54:07 2003

seq74-plus-73.rapb

Page 4

SEQ ID NO 13
LENGTH: 31
TYPE: PRT
ORGANISM: Bos taurus
US-09-932-161-13

Query Match 93.6%; Score 146; DB 10; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXRGIRPVGRF 31
DB 1 SRAHSHMEIRTPDINPAMYAGRGIRPVGRF 31

RESULT 11
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 93.6%; Score 146; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXRGIRPVGRF 31
DB 1 SRAHSHMEIRTPDINPAMYAGRGIRPVGRF 31

RESULT 12
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 40
LENGTH: 32
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-40

Query Match 93.6%; Score 146; DB 12; Length 32;
Best Local Similarity 83.9%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXRGIRPVGRF 31
DB 1 SRAHSHMEIRTPDINPAMYAGRGIRPVGRF 31

RESULT 13
US-10-044-592-41
Sequence 41, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 41
LENGTH: 33
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-41

Query Match 93.6%; Score 146; DB 12; Length 33;
Best Local Similarity 83.9%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXRGIRPVGRF 31
DB 1 SRAHSHMEIRTPDINPAMYAGRGIRPVGRF 31

RESULT 14
US-10-044-592-28
Sequence 28, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 28
LENGTH: 98
TYPE: PRT
ORGANISM: Murine
US-10-044-592-28

Query Match 93.6%; Score 146; DB 12; Length 98;
Best Local Similarity 83.9%; Pred. No. 6.4e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

